ENHANCING PROTEIN SECRETION IN *ESCHERICHIA COLI* BY CODON ENGINEERING VIA TRANSLATION OPTIMIZATION AND GENOME SEQUENCE ANALYSIS OF A HYPERSECRETER MUTANT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Prateek Gupta

May 2009

ENHANCING PROTEIN SECRETION IN *ESCHERICHIA COLI* BY CODON ENGINEERING VIA TRANSLATION OPTIMIZATION AND GENOME SEQUENCE ANALYSIS OF A HYPERSECRETER MUTANT

Prateek Gupta, Ph. D.

Cornell University 2009

Escherichia coli is a common host for recombinant protein production for biotechnology applications. Secretion of recombinant proteins to the extracellular and periplasmic space has the potential to reduce protein aggregation and to simplify downstream purification. A directed mutagenesis approach (specifically changing abundant codons to synonymous rare codons in a specific region) previously resulted in an eight-fold improvement in active hemolysin (HlyA) secretion. Also, synonymous codon substitutions have been shown to alter protein folding and function; however this mechanism is not well understood. In the first part of this study, a series of experiments have been described to study the effect of synonymous rare codon clusters on protein folding and secretion via multiple pathways in *E. coli*. Significant improvement in extracellular and periplasmic secretion of various recombinant proteins was observed by synonymous rare codon engineering. The analyses also revealed that synonymous rare codon cluster at specific sites of the target gene can alter polypeptide folding and activity by modulating the interactions of polypeptide with molecular chaperones. The study provides an experimental toolkit to enhance recombinant protein secretion in *E. coli* and offer insights into the effect of silent mutations on protein folding. The second part of the thesis includes genome sequence analysis of a derivative hypersecreter *E. coli* strain (B41), created previously by

random mutagenesis, and the parent strain using a 'next-generation' sequencing technology. Mutational profiling revealed a single nucleotide polymorphism (G → T) in the B41 genome, which results in premature translation termination of a transcription factor, RutR. Comparative mRNA expression analysis revealed that absence of RutR coordinates a decrease in the expression of tRNA-synthetases and some amino acid transporter genes, suggesting that the absence of RutR may result in slower translation rate. The work presents a single gene target to enhance extracellular secretion via the Type-I pathway and highlight the potential of new high-throughput massively parallel sequencing technologies to characterize selected mutants for strain improvement.

BIOGRAPHICAL SKETCH

The author was born in Ajmer, a beautiful city in the western part of India. He grew up in North India, near Delhi, where he completed his pre-undergraduate education and also met his future wife in high school. He then moved to the Indian Institute of Technology, Delhi to pursue undergraduate education in Biochemical Engineering and Biotechnology. Early on in his studies, he had an opportunity to interact with a professor in the Department of Chemistry who introduced him to the fascinating world of protein folding. Intrigued by the folding problem, the author spent a summer, after his sophomore year, at the Tata Institute of Fundamental Research (Bombay, India) learning various molecular biology and biophysical characterization techniques. He also spent a summer, after his junior year, in Germany learning NMR methods to solve secondary structure of proteins. The combined experience motivated him to pursue research after finishing his undergraduate degree. The author decided to come to Chemical Engineering Department at Cornell University in the fall of 2004 and was fortunate to work in Prof. Kelvin Lee's lab. He had a memorable time in Ithaca with his group mates during ho-ho's and with his friends playing cricket and visiting places of natural beauty. He stayed in Ithaca for three years working on devising ways to increase recombinant protein secretion in *E. coli* and then moved with his group to Delaware. Meanwhile he also got married to his high school sweetheart. Among the most memorable moments, the author remembers time spent with the undergraduate students in the lab when they got excited seeing their first results.

The author likes to play cricket and squash, read and cook in his time away from the lab.

Dedicated to my parents

ACKNOWLEDGMENTS

I am very fortunate to have been advised by Prof. Kelvin H. Lee in the course of my graduate studies. Kelvin has been not only a never-ending source of inspiration, but also a remarkably astute coach. He has taught me invaluable lessons and I can only hope that his technique, taste, and attitude continue to influence my work. I would also like to thank my committee members, Prof. Michael L. Shuler and Prof. Matthew P. DeLisa, for all their helpful advice, guidance and encouragement. Many useful discussions with Prof. DeLisa helped me to define the course of this project. I would also like to acknowledge my previous mentors, Prof. B. Jayaram, Prof. Shobhona Sharma, Prof. Christian Freund and Prof. Tapan Chaudhuri, who encouraged and motivated me to pursue research.

Several others have been part of this endeavor in their own ways. I would like to thank those who helped in this research. Leila Choe has been instrumental in collecting all the mass spectrometry data in this study. She has been a great support system in the lab and has been an extremely patient listener in our conversations over the years. I would like to thank previous Lee group members Erin Finehout, Chen Li, Kunal Agarwal, Bob Kuczenski, Mark D'Ascenzo, Dacheng Ren, Brenda Werner and Heather Roman for their help, support and initial mentoring. Special thanks to Bob for several discussions and working hard with me on the secretion model. I would like to acknowledge current Lee group members Pei Yu-Liao, Jeff Swanberg, Gilda Shayan, Yong Choi, Anup Agarwal, Stephanie Hammond and Jeff Foltz for their friendship and support. Pei has been an excellent colleague right from the first year and provided help and suggestions with many experiments, Jeff performed the genome sequencing analysis, Anup helped in some of the experiments towards the end, and Stephanie

proof read my papers. All the group members have been an amazing bunch of people to work with and I have learnt a lot from them during several scientific and non-scientific discussions we have had over the years. I would also like to thank Lydia Contreras, Adam Fisher, Matt Marrichi and Ritsdeliz Rodriguez in the Delisa laboratory for sharing the strains and plasmids and helping with some of the experiments. I also thank Dr. Peter Schweitzer and James VanEe for performing the genome sequencing experiments at the Biotechnology Research Center, Cornell University. I also had the privilege to mentor 4 undergraduate students (Sarah Mangan, Gautham Sreedharan, Emily Reasor and Divyanshu Pandey) and I thank them for bearing with me and working hard in the lab. I have learnt a lot from them.

I would like to thank all my friends at Cornell for their support and friendship. Hitesh Arora, Manish Sharma, Vikram Singh, Chandrani Roy Chowdhury, Gaurav Charaya, Bettina Susan John, Chris Orilall, Joe Goose and Abhishek Dube have been great friends and my support system in Ithaca. Also, I would like to acknowledge my friends outside Ithaca, Surbhi Jain, Yogesh Sharma, Kunal Gangwal, Shailendra Singh, and Amarjeet Singh who have been there with me to share the excitements and hardships of graduate work.

I am very grateful to my parents and family for all the support, caring and encouragement. My brother and sister in-law have been very considerate and motivated me throughout the course of my PhD. I am grateful to my parents in-law and brother in-law for their love and support, they have been exceptionally supportive. My nephew Atharv has been the stimulation behind many interesting anecdotes. Finally, I thank my wife Deepti for all that she is, caring, encouraging and supporting me when the going was tough.

I gratefully acknowledge the funding sources for this project from the National Science Foundation and New York State Foundation for Science, Technology and Innovation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

Recombinant DNA technology is used for the production of therapeutic or industrially relevant proteins in bacterial, yeast and other cell types. The annual global market for biopharmaceuticals is estimated at more than $33 billion and products derived from *Escherichia coli* represent nearly 30 % all the biopharmaceuticals made in the US and EU [1]. As listed in Table 1.1, a number of important biopharmaceuticals are produced in *E. coli*, including vaccines, interleukins, and hormones [1]. Thus, improvements in the production of heterologous protein in *E. coli* can have enormous economic impact. *E. coli* has become the host organism of choice for the expression of many heterologous genes, primarily because of the ease with which the organism may be genetically manipulated. Compared to other organisms, recombinant protein production in *E. coli* is simpler due to high growth rate, simple nutritional requirements, genetic stability, and ease of product purification. *E. coli* can be used in a wide range of applications, particularly where post-translational modifications are not required for the functions of target proteins. Furthermore, this organism has the ability to overexpress recombinant proteins to at least 20% of total cellular protein and to export them from the inside of the cell (cytoplasm) to the outside (periplasm or extracellular medium) [2].

Expression of heterologous proteins in *E. coli* can be divided into two major categories based on the final destination of the protein being processed. Heterologous proteins may be expressed to accumulate in the cytoplasm (intracellular expression), or the protein may be exported into the periplasmic space or culture media via

Table 1.1. Biopharmaceuticals that are produced in *Escherichia coli* and have been approved by FDA in 2006. Six categories of products and their therapeutic indication are shown.

Product	Therapeutic Indication
Recombinant anticoagulants (Tissue plasminogen factor)	Acute myocardial infarction
Recombinant hormones (Insulin, Human growth hormone)	Diabetes mellitus, growth hormone deficiency in children, Osteoporosis
Recombinant hematopoietic growth factors (GM-CSF)	Autologous bone marrow transplantation
Recombinant interferons and interleukins (Interferon-α, Interferon-β, Interleukin-1, Interleukin-11, Interleukin-12)	Hepatitis C, Multiple sclerosis, Rheumatoid arthritis, Chemotherapy induced thrombocytopenia
Recombinant vaccines	Lyme disease vaccine
Monoclonal Antibody based products (Recombinant tumor Necrosis Factor- α)	Adjunct to surgery for subsequent tumor removal

translocation pathways (secretion). Intracellular expression of proteins in the cytoplasm often results in the formation of inclusion bodies, which are unwanted protein aggregates [3]. These aggregates are not functional and are often difficult to process, thereby increasing production cost and reducing yield. Secretion of heterologous proteins in the periplasm can simplify purification processes, but often results in production bottlenecks because the translocation pathways are difficult to sustain at high levels of transport and the mechanisms of export are not well understood. The periplasm of *E. coli* contains fewer protein species and proteases than the cytosol, reducing the complexity of the initial mixture of contaminants. However inclusion bodies may also form in the periplasm as a result of high concentration of recombinant proteins [4].

From a biotechnology perspective, heterologous protein secretion into the extracellular medium is desirable because concentration of the protein of interest remains low, thus minimizing protein aggregation. Additionally, the extracellular environment can be controlled to provide optimal osmolarity and pH for protein folding and stability. However, there have been few successful attempts to secrete proteins into the extracellular medium, some of which have been reported in [2] and [5-9]. Efforts to alleviate bottlenecks in secretion pathways via metabolic engineering do not often produce the desired phenotype, primarily because of the low efficiency and high specificity of most secretion systems and an incomplete understanding of their mechanisms. While *E. coli* exports several types of proteins to the periplasm, secretion pathways to the extracellular medium are often specific for only a few proteins. Understanding and controlling translocation mechanisms enables better control over the secretion systems for the production of recombinant proteins.

Previous studies have established that the protein synthesis rate is an important consideration for secretion efficiency [8, 10]. Of the number of factors that can affect protein synthesis rate; plasmid copy number, translation initiation region, and differences in codon usage have been studied in greater details in the context of secretion. Plasmid copy number can affect the number of available ribosomes for translation and studies have shown that *E. coli* transformed with plasmids with different copy numbers exhibited altered export of human proinsulin to the periplasm [11]. It was found that the use of plasmids of moderate copy number (15-60) versus low copy number (11) increased export, but this study did not include a high copy number plasmid. Studies on the export of recombinant proteins to the periplasm [10] revealed that a change in the sequence of the translation initiation region (TIR) affects the rate of translocation by the Sec pathway. However, an increase in protein expression levels brought about by differences in the strength of TIR did not uniformly increase the amount of protein secreted. Rather, the optimal TIR (and presumably, expression level) varied for each protein tested [10]. The investigators speculated that the change to the first few codons in the gene sequence that were made might affect the structure of the mRNA resulting in enhanced binding and translation by the smaller ribosomal subunit. Nevertheless, this study demonstrated that an optimum translational level exists to achieve high-level secretion of each heterologous protein, and outside of this optimum level, secretion levels drop off significantly.

Differences in codon usage between the source organism and the production host have also been observed to affect the translation rate of recombinant proteins because codon usage directly correlates with the abundance of codon-specific tRNAs available within the cell [12]. Generally, a synonymous substitution of the codons in the gene of interest to more commonly used codons in the host cell alleviates this problem [13].

Recent studies have demonstrated that a decrease in the translation rate of HlyA leads to an enhanced secretion capability via Type-1 secretion machinery [8]. In particular an *E. coli* W3110 strain, designated as hly-slow, was engineered to have a predicted 37% decrease in protein production rate. This strain showed an eight-fold improvement in hemolysin secretion as compared to the control strain, designated as Hly-parent. The relative decrease in the translation rate was achieved by changing five of the adjacent codons to rare codons in a specific region of *hly*A gene but without altering the amino acid sequence. Rare codons are defined as those codons whose corresponding tRNA concentration is less than 1% of the total tRNA concentration (as tabulated in [24]). The use of rare codons in a gene has been shown to reduce the translation rate because the number of available aminoacyl-tRNAs is limited [14].

Studies have also shown that synonymous codon substitutions can change protein structure and function, indicating that protein structure is DNA sequence dependent. Synonymous codon substitutions that change codon usage frequencies from infrequent to frequent usage in regions of slow mRNA translation can deleteriously affect enzyme activity [15]. Conversely, synonymous substitutions that introduce rare codons into regions predicted to contain high frequency codons show altered substrate specificities [16]. Thus, contrary to conventional thinking, synonymous codon substitutions may not always be silent; changing codon usage frequency affects protein structure and function, and the frequency with which codons are used imparts vital information for the development of secondary and tertiary protein structure. Species-specific disparities in codon usage are frequently cited as the cause for failures in recombinant gene expression by heterologous expression hosts. Such failures include lack of expression, or the expression of protein that is non-functional or insoluble, or protein that is truncated because of proteolysis or premature

termination of translation [17–19]. All but the last of these failures are attributable to misfolded protein. In *E. coli*, as well as eukaryotic species, nascent proteins fold co-translationally within the ribosomal tunnel, which is both a protective environment within which secondary structure begins to form [20–22], and a dynamic environment that influences nascent protein structure [23–26]. Within the ribosomal tunnel, subtle variations in the rate of mRNA translation may play a key role in developing secondary structure in the nascent protein. Translation is not a steady state process, rather it occurs in pulses, as can be observed from ribosomal pausing [27] and even ribosome stacking, on specific stretches of mRNA [28]; these temporal changes in translational rate have been shown to depend on relative tRNA levels [29].

tRNA isoacceptor abundance and isoacceptor usage frequencies are directly related for naturally occurring proteins from *E. coli* as well as from other organisms [30–33], and there is evidence that protein secondary structure is related to tRNA usage frequencies [34], although this concept is controversial [35]. Comparative analysis of *E. coli* gene sequences and their respective protein structures show that amino acid sequences encoded by more frequently used codons are associated with highly ordered structural elements such as alpha helices, while sequences containing clusters of less frequently used codons tend to be associated with the protein domain boundaries (link/end segments) that separate such elements [36]. That analysis also showed that the link/end segments tend to be populated with amino acids that have bulky hydrophobic side chains or side chains that can hydrogen bond to the peptide backbone. When such residues appear in link/end segments, they tend to be encoded by infrequently-used codons. Therefore, the positioning of clusters of relatively high and low abundance codons on mRNA transcripts may be a purposeful rather than a random occurrence [37]. The idea that link/end segments, which separate elements of higher order protein

structure, are encoded by clusters of low-usage frequency codons leads to the hypothesis that slow translational progression (i.e., ‘‘pausing’’) through such regions of mRNA would allow the preceding nascent structural element to fold, at least partially, within the environment of the ribosomal tunnel prior to initiation of synthesis of the next structural element [38]. Such a temporal control mechanism would minimize the interaction between partially folded nascent polypeptides in the cytosol, an event which can lead to degradation, or aggregation and precipitation.

From a system’s perspective, the study of secretion systems has focused primarily on the proteins involved in secretion machinery, while less is known about the regulation of associated pathways, which may be crucial in alleviating bottlenecks in the process. The availability of genome sequence information has ushered in a new era of biology focused at the organism-wide level. Application of genomic and proteomic technologies has generated a lot of interest in the bioprocess development, because these tools provide molecular level details and aid in exploration of cellular functions to aid strain improvement [39]. Powerful techniques, such as cDNA microarrays and two-dimensional electrophoresis coupled with mass spectrometry, can provide interesting gene and protein expression changes which can give important clues to understand cellular phenotype. However, these tools do not provide the sequence information of the changes in the organism’s genome which likely induce the observed expression changes.

DNAsequencing is a central technology in our understanding of the molecular basis of many fundamental problems of biology. The Human Genome Project and the resequencing of selected regions of the human genome in disease association studies have contributed to a refined understanding of the molecular basis of many

diseasesThe so-called ‘next-generation’ sequencing technologies like Illumina® Genome Analyzer, 454® Roche, Applied Biosystems SOLID™ offer dramatic cost reductions and throughput increases and have revolutionized the drug discovery and pharmaceutical development process. These new technologies are being used extensively to understand disease association, drug resistance, biomarkers, and develop molecular diagnostics [40-43]. From a biotechnology standpoint, the use of these technologies can accelerate strain improvement because traditional genetic tools cannot easily identify the mutations related to interesting phenotypes, generated by genetic mutational studies, adaptive evolution, and phenotypic screening. Moreover, these technologies can be used in tandem with “-omics” tools to gain insights into specific gene functions and provide interesting targets for metabolic engineering.

1.2 Project Goals

The overarching objective of this project is to understand and develop methods to engineer an enhanced capability of *E. coli* cells to secrete active recombinant proteins. The main sub-goals of this work are as follows:

1. Understand the effect of synonymous mutations on the secretion of recombinant proteins via the Type-I pathway and other secretion pathways. This goal is based on previous observations [8,45].
2. Utilize the secretion pathways and their quality control mechanisms to explore the effect of synonymous rare codon clusters on protein folding and interactions of polypeptide with molecular chaperones. This goal is based on the idea that slower translation progression through the rare codon cluster can affect the folding of the preceding nascent structural element or its interaction with accessory proteins, within the environment of the ribosome tunnel [44].

3. Characterize a selected hypersecreter *E. coli* mutant strain at the genome sequence level using new genome sequencing technologies. This goal is based on our previous observation [8] and aims to discover novel metabolic targets for hypersecretion of recombinant proteins in *E. coli*.

1.3 Scope of Work

This dissertation first provides some background information about different systems for export to the periplasm and extracellular medium employed by *E. coli* (Chapter 2). These diverse pathways are important for both biotechnology applications and the study of infectious diseases. Chapter 3 presents a general review of the applications of genomics and proteomics research in the process development research, adapted from [39]. This chapter discusses the application of genome-wide expression profiling tools in the design and optimization of bioprocesses. Chapter 4, adapted from [45], investigates the biological basis for the observed phenomenon that the translation rate of the HlyA protein may be related to the ability to secrete higher levels of HlyA via the Type-I pathway. A detailed comparative analysis between a hypersecreter mutant strain (hly-slow) and a control strain (hly-parent) has been provided. Chapter 5 explores the utilization of different secretion pathways in *E. coli* to study the effect of synonymous rare codon cluster on polypeptide folding and secretion. This chapter illustrates how the introduction of rare codons in a specific stretch of the target gene affects the interaction of the polypeptide with molecular chaperones and highlights the benefit of synonymous codon engineering in improving protein secretion via multiple secretion pathways for bioprocess applications. Chapter 6 discusses application of a 'next-generation' genome sequencing technology for strain improvement in *E. coli*. Specifically, the study presents the discovery of a single nucleotide polymorphism in a derivative hypersecreter *E. coli* strain which can enhance the extracellular secretion of

recombinant proteins via the Type-I pathway. Chapter 7 summarizes the results and conclusions of this work and recommends areas of future work.

REFERENCES

1. Walsh G. 2006. Biopharmaceutical benchmarks - 2006. *Nat. Biotechnol.* **24**: 769-776.

2. Blight MA, Holland IB. 1994. Heterologous protein secretion and the versatile *Escherichia coli* haemolysin translocator. *Trend. Biotechnol*. **12**: 450- 455.

3. Villaverde A, Carrio MM. 2003. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol. Lett.* **25**: 1385-1395.

4. Georgiou G, Telford JN, Shuler ML, Wilson DB. 1986. Localization of inclusion bodies in *Escherichia coli* overproducing beta-lactamase or alkaline phosphatase. *Appl. Environ. Microbiol.* **52**: 1157-1161.

5. Fu J, Wilson DB, Shuler ML. 1993. Continuous, high level production and excretion of a plasmid-encoded protein by *Escherichia coli* in a two stage chemostat. *Biotechnol. Bioeng*. **41**: 937- 946.

6. Zhang G, Brokx S, Weiner JH. 2006. Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in *Escherichia coli*. *Nat Biotechnol.* **24**: 100-4.

7. Li YY, Chen CX, von Specht BU, Hahn HP. 2002. Cloning and hemolysin mediated secretory expression of a codon-optimized synthetic human interleukin-6 gene in *Escherichia coli*. *Prot. Expr. Purif.* **25**: 437- 447.

8. Lee PS, Lee KH. 2005. Engineering HlyA Hypersecretion in *Escherichia coli* Based on Proteomic and Microarray Analyses. *Biotechnol. Bioeng* . **89**: 195-205.

9. Sugamata Y, Shiba T. 2005. Improved secretory production of recombinant proteins by random mutagenesis of hlyB, an alpha-hemolysin transporter from *Escherichia coli*. *Appl. Environ. Microbiol.* **71**: 656-62.

10. Simmons LC, Yansura DG. 1996. Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. *Nat. Biotechnol.* **14**: 629-634.

11. Mergulhao FJ, Monteiro GA, Larsson G, Sanden AM, Farewell A, Nystrom T, Cabral JM, Taipa MA. 2003. Medium and copy number effects on the secretion of human proinsulin in *Escherichia coli* using the universal stress promoters uspA and uspB. *Appl. Microbiol. Biotechnol.* **61**:495-501.

12. Ikemura T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes:

a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* **151**: 389-409.

13. Williams DP, Regier D, Akiyoshi D, Genbauffe F, Murphy JR. 1988. Design, synthesis and expression of a human interleukin-2 gene incorporating the codon usage bias found in highly expressed *Escherichia coli* genes. *Nucleic Acids Res.* **16**: 10453-67.

14. Ikemura T. 1981a. Correlation between the abundance of *E. coli* transfer RNAs and the occurrence of the respective codon in the protein genes. *J. Mol. Biol.* **146**: 1-21.

15. Komar AA, Lesnik T, Reiss C. 1999. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett* **462**: 387-391.

16. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, et al. 2007. A ''silent'' polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**: 525-528.

17. Adzhubei AA, Adzhubei IA, Krasheninnikov IA, Neidle S. 1996. Non-random usage of 'degenerate' codons is related to protein three-dimensional structure. *FEBS Lett* **399**: 78-82.

18. Kurland C, Gallant J. 1996. Errors of heterologous protein expression. *Curr. Opin. Biotechnol.* **7**: 489-493.

19. Lindsley D, Gallant J, Guarneros G. 2003. Ribosome bypassing elicited by tRNA depletion. *Mol. Microbiol.* **48**: 1267-1274.

20. Kleizen B, van Vlijmen T, de Jonge HR, Braakman I. 2005. Folding of CFTR is predominantly cotranslational. *Mol. Cell* **20**: 277-287.

21. Kramer G, Ramachandiran V, Hardesty B. 2001. Cotranslational folding–omnia mea mecum porto? Int J Biochem *Cell Biol.* **33**: 541-553.

22. Svetlov MS, Kommer A, Kolb VA, Spirin AS. 2006. Effective cotranslational folding of firefly luciferase without chaperones of the Hsp70 family. *Protein Sci*. **15**: 242-247.

23. Etchells SA, Hartl FU. 2004. The dynamic tunnel. *Nat. Struct. Mol. Biol.* **11**: 391-392.

24. Woolhead CA, McCormick PJ, Johnson AE. 2004. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116**: 725-736.

25. Baram D, Yonath A. 2005. From peptide-bond formation to cotranslational folding: dynamic, regulatory and evolutionary aspects. *FEBS Lett.* **579**: 948-954.

26. Berisio R, Schluenzen F, Harms J, Bashan A, Auerbach T, et al. 2003. Structural insight into the role of the ribosomal tunnel in cellular regulation. *Nat. Struct. Biol.* **10**: 366–370.

27. Purvis IJ, Bettany AJ, Santiago TC, Coggins JR, Duncan K, et al. 1987. The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. *J. Mol. Biol.* **193**: 413-417.

28. Wolin SL, Walter P. 1988. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J.* **7**: 3559-3569.

29. Varenne S, Buc J, Lloubes R, Lazdunski C. 1984. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J. Mol. Biol.* **180**: 549-576.

30. Bulmer M. 1987. Coevolution of codon usage and transfer RNA abundance. *Nature* **325**: 728–730.

31. Gouy M, Gautier C. 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* **10**: 7055-7074.

32. Ikemura T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* **151**: 389-409.

33. Ikemura T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* **146**: 1-21.

34. Thanaraj TA, Argos P. 1996. Ribosome-mediated translational pause and protein domain organization. *Protein Sci.* **5**: 1594-1612.

35. Oresic M, Shalloway D. 1998. Specific correlations between relative synonymous codon usage and protein secondary structure. *J. Mol. Biol.* **281**: 31-48.

36. Thanaraj TA, Argos P. 1996. Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci.* **5**: 1973-1983.

37. Phoenix DA, Korotkov E. 1997. Evidence of rare codon clusters within *Escherichia coli* coding regions. *FEMS Microbiol. Lett.* **155**: 63-66.

38. Angov E, Hillier CJ, Kincaid RL, Lyon JA. 2008. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS One*. **3**: e2189.

39. Gupta P. Lee K.H. 2008. Genomics and proteomics in process development: opportunities and challenges. *Trends Biotechnol.* **25**: 324-330.

40. Ley, T.J. *et al.* 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**: 66-72.

41. Thomas, R.K. *et al.* 2007. High-throughput oncogene mutation profiling in human cancer. *Nat. Genet.* **39**: 347–351.

42. Crossman, L. *et al.* 2008. The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biol.* **9**, R74.

43. Dahl, F. *et al.* 2007. Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc. Natl. Acad. Sci. USA* **104**: 9387–9392.

44. Frydman J. 2001.Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu. Rev. Biochem.* **70**: 603-647.

45. Gupta P. Lee K.H. 2008. Silent mutations result in HlyA hypersecretion by reducing intracellular HlyA protein aggregates. *Biotechnol. Bioeng.* **101**: 967-974.

CHAPTER 2

SECRETION PATHWAYS IN *ESCHERICHIA COLI*

2.1 Preface

This chapter provides an overview of the Type-I and Type-II secretory pathways that have evolved to translocate proteins across membranes in *Escherichia coli*. The secretory production of recombinant proteins by *E. coli* has several advantages over intracellular production. In most cases, targeting protein to the periplasmic space or to the culture medium facilitates downstream processing, folding, and *in vivo* stability, enabling the production of soluble and biologically active proteins at a reduced process cost. Systems for the export of proteins to the periplasm and those for secretion to the extracellular milieu are described.

2.2 Introduction

Most bacteria secrete proteins such as degradative enzymes, toxins, and other pathogenicity factors into the extracellular environment [1]. In Gram-negative bacteria, secreted proteins have to cross the two membranes of the cell envelope, which differ substantially in both composition and function [2].

The type I, II, III, IV, and V secretion pathways are widespread among Gram-negative bacteria and their mechanisms differ significantly. Despite these differences, the systems have, in common, a need to recognize specifically their cognate substrates and promote secretion without compromising the barrier function of the cell envelope [3]. This chapter discusses the Type-I secretion pathway for extracellular secretion and the Sec, SRP and Tat mechanisms for periplasmic secretion, which are used most commonly for recombinant protein secretion in *E. coli*. A brief discussion about Type-

II extracellular secretion is also provided, which involves a two step process. The type III secretion pathway is characteristic of several pathogenic Gram-negative bacteria and has been reviewed [4]. Type IV secretion comprises those pathways usually found in bacterial conjugation systems [5] and has been reviewed [6]. The type V mechanism includes the autotransporter and the two-partner secretion systems [5], and has been reviewed [7].

Bacteria often sequester highly expressed expressed recombinant protein in inclusion bodies, which are aggregates of precipitated protein. These inclusion bodies can be isolated and the protein can be refolded, but this recovery often requires additional steps in downstream processing that can be costly. Often, purification of proteins of interest requires cell disruption, increasing the number of contaminant species that must be removed and exposing the desired product to more proteases [8]. Secretion of recombinant proteins to the culture medium or periplasm of *E. coli* has several advantages over intracellular production. These advantages include simplified downstream processing, enhanced biological activity, higher product stability and solubility, and N-terminal authenticity of the expressed peptide. The periplasm of *E. coli* is an oxidizing environment, which aids in the formation of disulfide bonds [8]. It also contains fewer protein species and proteases, reducing the complexity of the initial mixture of contaminants. However, overexpression of recombinant proteins can also cause formation of inclusion bodies in the periplasm [9], likely due to the lack of ATP-dependent chaperones in this compartment. Secretion to the extracellular medium is highly desirable because the concentrations of protein of interest will remain low, reducing the risk of aggregation. Further, most laboratory strains of *E. coli* do not secrete any proteins to the extracellular medium [10]. Thus, problems associated with protease degradation are mitigated, and downstream purification of the

secreted protein is greatly facilitated. The extracellular medium can also be supplemented with cofactors or adjusted to optimal pH or osmolarity conditions to facilitate or to stabilize folding for the protein of interest. Understanding and controlling protein translocation mechanisms can greatly facilitate recombinant protein production.

2.3 Type I secretion

Type I secretion systems (T1SS) are characterized by the ATP-binding cassette (ABC) superfamily of proteins, which are involved in many transport functions in eukaryotes and prokaryotes. In Gram-positive bacteria, these systems participate in drug efflux and in toxin and cytolysin secretion. Gram-negative bacteria also use ABC transport systems for uptake of iron siderophores, carbohydrates, amino acids, and oligopeptides [11]. ABC protein transporters are also widespread in eukaryotes, where they have been found to play a critical role in resistance of cancers to chemotherapy drugs and in the genetic mutation that causes cystic fibrosis [12]. The ABC protein transporter in Gram-negative bacteria translocates peptides across both the inner and outer membranes in a single step [13]. These transporters have evolved to transport specific target peptides, which include toxins, proteases, lipases, and hemoproteins [11]. Despite their specificity, they are widespread among Gram-negative bacteria and are highly conserved. The α-hemolysin (HlyA) secretion pathway in pathogenic *E. coli* is the most extensively studied T1SS.

A signal sequence is required for protein translocation and is conserved for particular families of secreted proteins. Typically, the signal sequence is located at the C-terminus, although there is one notable exception in the export of the antibiotic peptide colicin V in *E. coli*, with a signal sequence located at the N-terminus [14]. In proteases

and lipases secreted by Type I systems, the C-terminus consists of several negatively charged amino acid residues followed by several hydrophobic ones and appears to form a stable α-helix [15]. Toxins, proteases, and lipases secreted by ABC transporters have repeating glycine and aspartic-rich (RTX) domains just upstream of the C-terminus signal sequence that are critical for secretion. It is hypothesized that these RTX regions act as internal chaperones and aid in distinguishing secretory from cytosolic proteins [16]. T1SS signal sequences have low sequence homology but the general motif is conserved in these families. Mutagenesis experiments of the sixty amino acid signal sequence of the toxin α-hemolysin in *E. coli* have shown that single mutations in the region do not decrease export efficiency; however, combinations of mutations do, implying that tertiary structure may be important in signal sequence recognition [17]. Signal sequences are not cleaved at any point in the translocation process. Hence for certain biotechnology applications (production of therapeutic proteins), protease sites such as thrombin can be incorporated between the reporter protein and the signal sequence.

Gram-negative ABC protein-mediated transport systems are composed of three different secretion machinery proteins: the ABC protein, a membrane fusion protein (MFP) and an outer membrane pore protein. ABC proteins are highly conserved among many different species and functions. They usually consist of two inner transmembrane domains and two hydrophilic (cytosolic) ATP-binding domains [15]. These cytosolic domains have ATPase activity *in vitro* [11] and probably provide energy for the translocation reaction; however, it is unknown which exportation steps require ATP hydrolysis, or how the energy is transferred to the secretion process. Studies indicate that the ABC protein is responsible for substrate specificity in the transport system [17]. The ATP binding-domain of HlyB, the ABC protein involved in

secretion of α-hemolysin has recently been crystallized [18]. It consists of a catalytic domain that is highly conserved among ABC transporters, and a signaling domain specific to each transporter. The catalytic domain is comprised of two β-sheets and seven α-helices and is responsible for hydrolysis of ATP. The signaling domain consists of five α-helices and a highly variable C-loop, which is thought to interact with the transmembrane domains and to recognize the transport substrate. Experiments have shown that ATP hydrolysis is required for translocation but not for assembly of the translocation complex [17].

One of the membrane helper proteins is a member of the membrane fusion protein family (MFP), often involved in membrane transporters and antiporters. MFP genes are often adjacent to the genes encoding the ABC proteins [11]. Sequence predictions indicate that they are comprised of an N-terminal hydrophobic segment anchored in the inner membrane, a large hydrophilic region, probably located in the periplasm, and a C-terminus β-barrel structure that interacts with the outer membrane protein [15]. The MFP is thought to permit a localized fusion of the two membranes. Recent studies in the Hly pathway indicate that TolC trimers in the outer membrane, surrounded by several HlyD monomers which also surround HlyB in the inner membrane, forming a continuous sealed channel from the cytoplasm to the exterior, with a molecule of HlyA in transit (Fig. 2.1) [19]. The actual stoichiometry of HlyD with respect to other components remains unclear [19].

The other membrane protein forms a pore in the outer membrane and may be general or specific to the translocation pathway. Biochemical evidence suggests that the outer membrane protein exhibits no substrate specificity [11]. In *E. coli* the same outer membrane protein TolC is used for secretion of both colicin V-1 and α-hemolysin,

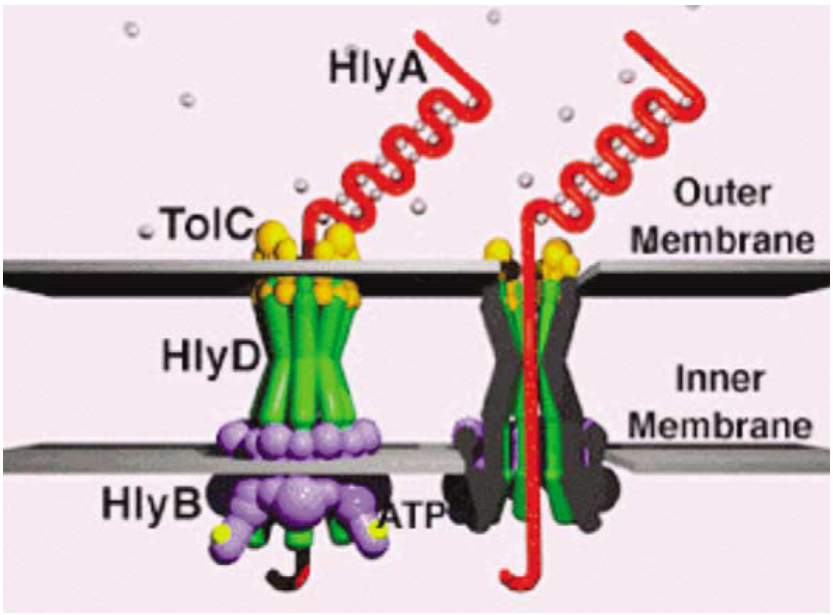


Figure 2.1. The Type1 translocator in *E. coli*. Cartoon representation of TolC trimers in the outer membrane, surrounded by several HlyD monomers which also surround HlyB in the inner membrane, forming a continuous sealed channel from the cytoplasm to the exterior, with a molecule of HlyA in transit. The actual stoichiometry of HlyD with respect to other components remains unclear. This figure is reproduced from [20].

while the ABC and MFP proteins are specific [14]. TolC also participates in efflux of small noxious molecules such as detergents and anti-bacterial agents. The crystal structure of TolC has been solved [21], and the homotrimer appears to form an outer membrane β-barrel component and a large α-helical barrel segment that may interact with HlyD in the periplasm [21,22]. The outer membrane portion resembles that of a porin, forming a pore with an interior diameter of 19.8 Å [23]. The periplasmic portion is unique in that it appears to be self-closing when not associated with the other secretion machinery proteins, preventing the free diffusion of hydrophilic molecules into or out of the cell [24].

Studies indicate that the substrate initiates the assembly of protein secretion machinery, and that subsequent interactions are very ordered [20]. Substrate affinity chromatography suggests that the substrate's C-terminus binds to the ABC protein, which then interacts with the MFP, which then associates with the outer membrane protein to effect secretion across both membranes [17]. Experiments show that the C-terminus of HlyD interacts with the periplasmic domain of HlyB and HlyA [19], while the outer membrane portion of HlyD interacts with TolC to form a pore through which HlyA is released [24]. TolC is endogenously expressed [25] and also participates in the import and export process of many smaller molecules. The *hly* operon includes the *hlyABD* genes, as well as the gene encoding HlyC, a small post-translational activator of HlyA that has been shown to play no role in translocation [25]. The Hly system has been used to create fusion proteins with C-terminal sequences that have been successfully secreted in *E. coli*. A diverse range of heterologous proteins, such as antibodies, hormones, and proteases, have been secreted [25,26]; however, much remains unclear about other factors which may affect the secretion machinery, and the system's specificity for secreted substrates. Early studies attempting secretion of

dihydrofolate reductase (DHFR) using the Hly system did not produce an active enzyme; however, a single amino acid change in DHFR fused to the signal sequence resulted in an active enzyme that was secreted to the media. Other recombinant proteins were found to be secretion incompetent in an active state using the Hly pathway.

Secretion in this pathway typically requires no chaperones [11], with the exception of the heme-binding metalloprotease HasA in *Serratia marcescens* that requires the SecB chaperone [26], possibly to prevent the protein from folding prior to translocation. The folding states of the protein at different points in the ABC transport pathway are as yet unclear, but folded HasA protein has been shown to be translocation-deficient, and may inhibit the translocation machinery by sequestering the HasABC transporter [27]. Experimental evidence indicates that SecB facilitates translocation of HasA by sequestering it in an unfolded state prior to translocation. Studies with single chain antibody fragments secreted using the Hly system indicate that folding and disulfide bonds formation occurs during the translocation process [28]. Our studies show that GFP can be secreted using the hemolysin pathway, but is inactive once translocated. As with the Sec pathway, GFP is probably maintained in an unfolded state prior to and during translocation, resulting in misfolded protein once outside the cell. The requirements for unfolded peptides prior to translocation and for unassisted folding in the extracellular medium limit the types of proteins that can be secreted in an active state using Type I secretion pathways.

Many research groups are studying the Type-I secretion pathway because of its efficiency and widespread occurrence in both prokaryotes and eukaryotes [29]. As mentioned previously, ABC export mechanisms are important in cancer and cystic

fibrosis research; they are also of interest to the biotechnology industry, where the simplicity of the ABC transporter system and relatively small number of required machinery proteins may facilitate implementation into recombinant protein processes. Other researchers are investigating the use of type I pathway systems for the production of attenuated vaccines [30].

2.4. Sec export pathway

The Sec system, or General Export Pathway (GEP) is responsible for export of most proteins in wild-type *E. coli* and most Gram-negative bacteria [11,31]. It utilizes approximately six proteins encoded by the *sec* gene cluster to translocate the substrate to the periplasm, although up to eight other helper proteins have been implicated. An N-terminal signal sequence [32] is recognized by chaperone proteins which prevent folding prior to translation; the protein is translocated into the periplasm in an unfolded state by the Sec system proteins [31,33]. Sec Signal sequences average 24 amino acids in length and contain three distinct regions: a positively charged N-terminus, a hydrophobic α-helical region, and a c-domain that contains the cleavage site [34]. During translation, the nascent preprotein (substrate with leader peptide sequence still attached) is bound by the chaperone SecB to prevent folding [33]. The specific interactions by which SecB identifies targets for export are unclear, as SecB has been shown to interact with nascent chains without a signal peptide. SecB delivers the preprotein to the site of translocation by binding to the membrane-bound SecA protein. It appears to bind directly to the signal sequence, and the preprotein is transferred from SecB to membrane-bound SecA. SecA can also dissociate from the membrane to bind to preproteins directly and transport them to the membrane [35]. SecY and SecE are inner membrane proteins that form a core translocation complex and are necessary for transport [33]. SecY appears to constitute the channel by which

the peptide is translocated, while the exact function of SecE is unknown [35]. Inclusion of SecG markedly increases secretion efficiency, and SecYEG is known to form a complex that binds SecA at the membrane. SecA, Y, E, and G are thought to possess a proofreading function, rejecting defective preproteins from export [36]. Proteins are translocated across the membrane in a stepwise fashion, powered by the ATPase function of SecA. The hydrolysis of ATP produces a conformational change in SecA, causing it to release and rebind the preprotein, thereby threading it through the SecYEG translocase complex (Fig. 2.2).

The membrane proteins SecD and SecF contain large periplasmic domains [31] and are known to complex with YajC and loosely bind to the SecYEG complex. SecD and F may also be involved in release of translocated proteins by facilitating folding or affecting the energized state of the membrane [36], as it has been shown that their presence is necessary for maintaining a ΔpH. Recent evidence also indicates that the SecDF/YajC complex may prevent backward sliding of the preprotein in the translocation channel [35]. Peptidases such as LspA and Lep in *E. coli* operate in the periplasm to cleave the leader peptides, yielding mature protein [33].

Studies using reporter proteins suggest that the Sec pathway in *E. coli* cannot translocate active green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* [36]. It is hypothesized that transport of unfolded GFP by the prokaryotic Sec export system renders the protein unable to fold correctly once released from the translocation channel. This observation highlights a limitation of the Sec pathway, namely that exported proteins must be able to fold in the chaperone-poor environment of the periplasm. To this end, several foldases including DsbA, which catalyzes the

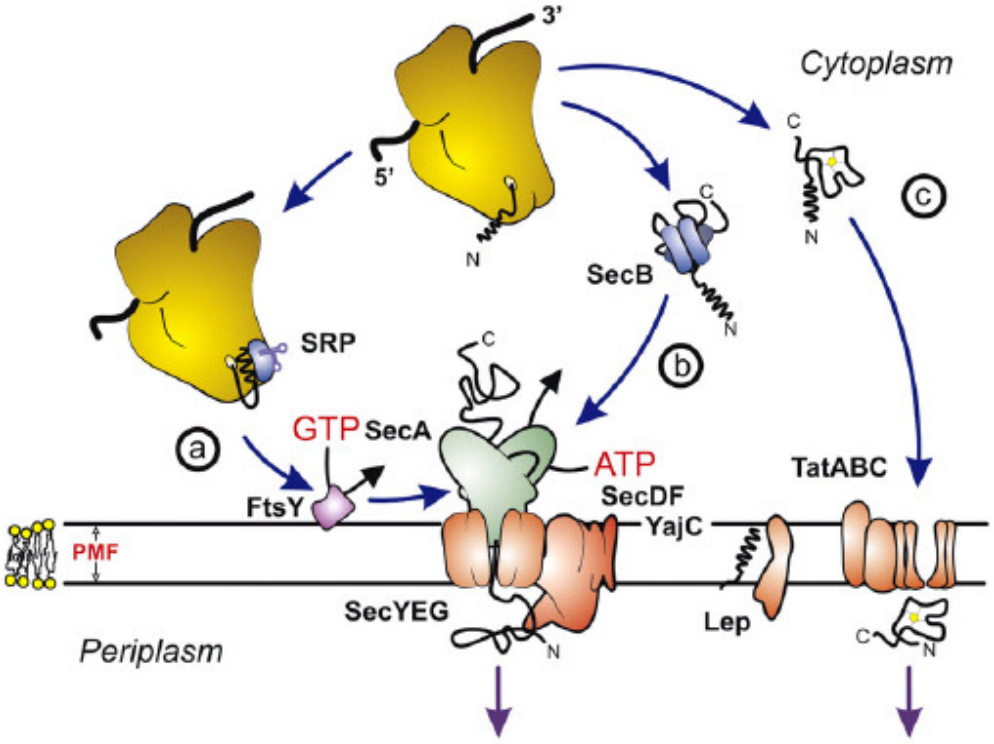


Figure 2.2. Schematic overview of the *E. coli* Sec, SRP and Tat translocases. (a) Co-translational translocation by the SRP pathway, (b) post-translational targeting routes and translocation of unfolded proteins by the Sec-translocase, and (c) translocation of folded precursor proteins by the Tat translocase. This figure is reproduced from [37].

formation of disulfide bonds [31], and Skp, which aids in folding and insertion of outer membrane proteins [38], are found in the periplasm. Genetic modifications of *E. coli* have allowed protein export of proteins to the periplasm via the Sec pathway and subsequent secretion to the extracellular medium due to increased outer membrane permeability [39,40].

2.5 SRP Pathway

The signal recognition particle (SRP) pathway is used by *E. coli* primarily for the targeting of inner membrane proteins [41]. This system has been exploited in the secretion of several recombinant proteins including Mtla–OmpA fusions [42], MalF–LacZ fusions [43], maltose binding protein, chloramphenicol acetyl transferase [44,45], and haemoglobin protease [46].

SRP recognizes substrates by the presence of a hydrophobic signal sequence (hence the name **s**ignal **r**ecognition **p**article). The presence of an N-terminal signal sequence with a highly hydrophobic core, combined with a lack of a trigger factor binding site [47], results in co-translational binding of the nascent chain to Ffh [48]. For a productive interaction between the preprotein and Ffh, 4.5S RNA is required [49]. It has been suggested [50] that the interaction between SRP and the signal sequence is dependent on the hydrophobicity of the nascent chain since preproteins with more hydrophobic signal sequences are translocated with higher efficiency. It has been shown [51] that SRP binds the ribosome at a site that overlaps the binding site of trigger factor. A discriminating process has been proposed in which SRP and trigger factor alternate in transient binding to the ribosome until a nascent peptide emerges. Depending on the characteristics of the nascent peptide, the binding of either SRP or trigger factor is stabilized, thus determining whether the peptide is targeted to the

membrane via the SRP pathway, or post-translationally by the SecB pathway [51].

FtsY is found both in the cytoplasm and at the membrane [49], and can interact with ribosomal nascent chain–SRP complexes in the cytosol (Fig. 2.2). Upon interaction with membrane lipids, the GTPase activities of FtsY and Ffh are stimulated, thus releasing the nascent chain to the translocation site [52]. This site may be the SecYEG translocon [53], although it has been demonstrated that membrane insertion can occur independently of SecYEG [54]. Insertion of transmembrane segments can occur in the absence of SecA [55] while translocation of large periplasmic loops is SecA-dependent [56].

For recombinant protein production, SRP targeting can be achieved by engineering the hydrophobicity of the signal sequence [57]. This is advantageous if the target protein folds too quickly in the cytoplasm, adopting a conformation incompatible with secretion by the SecB-dependent system [58].

2.6 Tat Pathway

The TAT pathway is capable of transporting folded proteins across the inner membrane [59] independently of ATP [60] using the transmembrane PMF [61]. In most cases, the substrates of this pathway are proteins that bind specific cofactors in the cytoplasm and are folded prior to export [62,63]. This system is related to the DpH-dependent protein import machinery of the plant chloroplast thylakoid membrane [64]. The TAT pathway has been used in the secretion of several recombinant proteins including antibody fragments [65], green fluorescent protein [66,67] and several others.

The main components of this translocation system in *E. coli* are TatA, TatB and TatC proteins. TatA has been proposed to form the transport channel [68], although TatAB complexes have also been implicated in that function [69]. TatB and TatC are proposed to form a 1:1 complex that may provide the initial binding site for preprotein docking [70-72]. It has also been proposed that the signal sequence is recognized by TatC and then transferred to TatB [73]. A mechanism for protein translocation by the TAT system was recently proposed [74]. In this study, the authors propose that the signal peptide is recognized by TatC, which is forming a complex with TatB. When signal peptide binding occurs, proton motive force promotes the association between the TatBC complex and TatA oligomers. The folded preprotein is then translocated by the TatA channel and the leader peptide is processed (Fig. 2.2). Following translocation, TatA dissociates from the TatBC complex [74]. It has also been reported that TatE can partially substitute TatA [75].

Like Sec signal peptides, TAT signal peptides are also composed of three regions: a positively charged region (n-region), a hydrophobic region (h-region), and a c-region that contains the cleavage site [76]. The average size of these signal peptides is approximately 38 amino acids, which is 14 amino acids longer than the average Sec leader peptide. Most of this additional length is due to an extended n-region. TAT signal peptides bear the N-terminal consensus motif S/T-R-R-X-F-L-K, where X is highly variable [76]. Although the presence of both arginine residues is not an obligatory requirement for transport, mutagenesis of one or both of these residues can affect membrane translocation [66,77]. The h-region of TAT signal peptides is usually less hydrophobic than that of Sec leader peptides. The c-region contains the cleavage site and shows a strong bias towards basic amino acid residues [75]. It has not been established whether these signal peptides are cleaved by signal peptidase I or by some

other protease [78].

Tat systems translocate folded proteins across biological membranes [79] and therefore it is believed that its physiological role is to extend the set of translocatable substrates to those that fold prior to translocation. This is the fundamental functional difference to the Sec system, which only translocates largely unfolded proteins [80]. It has been proposed that the Tat system is used only in cases where cytoplasmic folding excludes the use of the Sec system [81]. Tat substrates fold inside the cytoplasm and thus can use general chaperones for folding. Specific chaperones have been implicated in preventing translocase interactions prior to cofactor insertion and folding, leading in many cases to an oligomerization of the substrate protein [82]. This has been termed "proofreading" and used to describe a function that prevents targeting prior to folding [83]. Although general chaperones likely suffice to ensure folding of Tat substrates, an additional folding quality control at the Tat system may exist to ensure that only folded proteins can pass the translocase [84]. This selectivity at the Tat translocase can be observed with overexpressed folding-impaired artificial Tat substrates [84], but so far no evidence for a biological requirement of this quality control system at the Tat translocase has been demonstrated.

2.7 Type II secretion

The general secretory pathway is a two-step process for the extracellular secretion of proteins mediated by periplasmic translocation [3]. Three pathways can be used for secretion across the bacterial cytoplasmic membrane: the SecB-dependent pathway, the signal recognition particle (SRP), and the twin-arginine translocation (TAT) pathways. The second step (translocation across the outer membrane) involves specific protein machinery. Extracellular secretion by a type II mechanism constitutes the main

terminal branch (MTB) of the general secretory pathway. This step is complex and requires 12–16 proteins that constitute the secreton [85-88]. Although the functions of individual secreton components are not known, some roles have been attributed by comparative analysis with other secretons that are highly conserved among Gram-negative bacteria [87-89]. When exiting peptides reach the periplasm, they are thought to adopt tertiary and even quaternary structures in order to be recognised by the MTB components. Although it is known that proteins have to adopt secretion-competent conformations to proceed further [85-86], no secretion signal on the folded proteins has been identified [87].

2.8 Conclusion

Bacteria have evolved a diversity of mechanisms to transport proteins to the periplasm (via the Sec, SRP and Tat pathways), to the extracellular medium (via the Type-I and Type-II pathways), or directly into a target cell (via the Type-III and Type-IV pathways). The knowledge about the functional role of individual components in the secretion pathways can help us to understand disease progression related to various bacterial infections and may enable better understanding of secretion of therapeutic or vaccinating agents in gene therapy. Protein secretion in bacteria is also of great interest for applications in recombinant DNA technology and the advances in our understanding of specific pathway mechanisms can provide new tools and technologies to improve recombinant protein expression for the production of therapeutically relevant proteins.

REFERENCES

1. Fernandez LA, Berenguer J. 2000. Secretion and assembly of regular surface structures in Gram-negative bacteria. *FEMS Microbiol. Rev.* **24**: 21-44.

2. Koebnik R, Locher KP, Van Gelder P. 2000. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol*. **37**: 239-253.

3. Koster M, Bitter W, Tommassen J. 2000. Protein secretion mechanisms in Gram-negative bacteria. *Int J Med Microbiol.* **290**: 325-231.

4. Cornelis GR, Van Gijsegem F. 2000. Assembly and function of type III secretory systems. *Annu Rev Microbiol.* **54**: 735-774.

5. Pallen MJ, Chaudhuri RR, Henderson IR. 2003. Genomic analysis of secretion systems. *Curr Opin Microbiol.* **6**: 519-527.

6. Christie PJ. 2001. Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Mol Microbiol.* **40**: 294-305.

7. Jacob-Dubuisson F, Locht C, Antoine R. 2001. Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Mol Microbiol.* **40**: 306-313.

8. Georgiou G, Valax P. 1996. Expression of correctly folded proteins in *Escherichia coli*. *Curr Opin Biotechnol.* **7**: 190-197.

9. Georgiou G, Telford JN, Shuler ML, Wilson DB. 1986. Localization of inclusion bodies in *Escherichia coli* overproducing beta-lactamase or alkaline phosphatase. *Appl Environ Microbiol.* **52**: 1157-1161.

10. Pugsley AP, Francetic O. 1998. Protein secretion in *Escherichia coli* K-12: Dead or alive? *Cell Mol Life Sci.* **54**: 347-352.

11. Wandersman C. 1996. Protein export and secretion. *Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology. Neidhardt FC. Washington, ASM Press.

12. Higgins CF. 1992. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol*. **8**: 67-113.

13. Blight MA, Holland IB. 1994. Heterologous protein secretion and the versatile *Escherichia coli* haemolysin translocator. *Trends Biotechnol.* **12**: 450-455.

14. Zhong XT, Kolter R, Tai PC. 1996. Processing of colicin V-1, a secretable marker protein of a bacterial ATP binding cassette export system, requires

membrane integrity, energy, and cytosolic factors. *J Biol Chem.* **271**: 28057-28063.

15. Binet R, Letoffe S, Ghigo JM, Delepelaire P, Wandersman C. 1997. Protein secretion by Gram-negative bacterial ABC exporters - A review. *Gene* **192**: 7-11.

16. Duong F, Lazdunski A, Murgier M. 1996. Protein secretion by heterologous bacterial ABC-transporters: the C-terminus secretion signal of the secreted protein confers high recognition specificity. *Mol. Microbiol.* **21**: 459-70.

17. Thanabalu T, Koronakis E, Hughes C, Koronakis V. 1998. Substrate-induced assembly of a contiguous channel for protein export from *E. coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *EMBO J.* **17**: 6487-6496.

18. Schmitt L, Benabdelhak H, Blight MA, Holland BI, Stubbs MT. 2003. Crystal structure of the nucleotide-binding domain of the ABC-transporter haemolysin B: Identification of a variable region within ABC helical domains. *J Mol Biol.* **330**: 333-342.

19. Pimenta AL, Young J, Holland IB, Blight MA. 1999. Antibody analysis of the localisation, expression and stability of HlyD, the MFP component of the *E. coli* haemolysin translocator. *Mol Gen Genet.* **261**: 122-132.

20. Holland IB, Schmitt L, Young J. 2005. Type 1 protein secretion in bacteria, the ABC- transporter dependent pathway. *Mol. Membr. Biol.* **22**: 29-39.

21. Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C. 2000. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**: 914-9.

22. Lewis K. 2000. Translocases: a bacterial tunnel for drugs and proteins. *Curr. Biol.* **10**: R678-81.

23. Andersen C, Koronakis E, Bokma E, Eswaran J, Hymphreys D, Hughes C, Koronakis V. 2002. Transition to the open state of the ToIC periplasmic tunnel entrance. *Proc Natl Acad Sci USA* **99**: 11103-11108.

24. Andersen C, Hughes C, Koronakis V. 2000. Channel vision: Export and efflux through bacterial channel-tunnels. *EMBO Rep.* **1**: 313-8.

25. Holland IB, Kenny B, Steipe B, Pluckthun A. 1990. Secretion of heterologous proteins in *Escherichia coli*. *Methods Enzymol.* **182**: 132-143.

26. Delepelaire P, Wandersman C. 1998. The SecB chaperone is involved in the secretion of the *Serratia marcescens* HasA protein through an ABC transporter. *EMBO J.l* **17**: 936-944.

27. Debarbieux L, Wandersman C. 2001. Folded HasA inhibits its own secretion through its ABC exporter. *EMBO J.* **20**: 4657-4663.

28. Fernandez LA, de Lorenzo V. 2001. Formation of disulphide bonds during secretion of proteins through the periplasmic-independent type I pathway. *Mol. Microbiol.* **40**: 332-346.

29. Blight MA, Holland IB. 1994. Heterologous protein secretion and the versatile *Escherichia coli* haemolysin translocator. *Trends Biotechnol.***12**: 450-455.

30. Hahn HP, von Specht BU. 2003. Secretory delivery of recombinant proteins in attenuated *Salmonella* strains: potential and limitations of Type I protein transporters. *FEMS Immunol Med Microbiol.* **37**: 87-98.

31. Pugsley AP. 1993. The complete general secretory pathway in Gram-negative bacteria. *Microbiol Rev.* **57**: 50-108.

32. Blobel G, Dobberstein B. 1975. Transfer of proteins across membranes .1. Presence of proteolytically processed and unprocessed nascent immunoglobulin light-chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol.* **67**: 835-851.

33. Driessen AJM, Fekkes P and van der Wolk JPW. 1998. The Sec system. *Curr Opin Microbiol.* **1**: 216-222.

34. Cristobal S, de Gier JW, Nielsen H, von Heijne G. 1999. Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli*. *EMBO J.* **18**: 2982-2990.

35. Danese PN, Silhavy TJ. 1998. Targeting and assembly of periplasmic and outer membrane proteins in *Escherichia coli*. *Annu. Rev. Genet.* **32**: 59-94.

36. Feilmeier BJ, Iseminger G, Schroeder D, Webber H, Phillips GJ. 2000. Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli. J Bacteriol.* **182**: 4068-4076.

37. Natale P, Brüser T, Driessen AJM. 2008. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. *Biochim. Biophys. Acta*. **1778**: 1735-1756.

38. Schafer U, Beck K, Muller M. 1999. Skp, a molecular chaperone of Gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. *J Biol Chem.* **274**: 24567-24574.

39. Fu J, Wilson DB, Shuler ML. 1993. Continuous, high level production and excretion of a plasmid-encoded protein by *Escherichia coli* in a two-stage chemostat. *Biotechnol. Bioeng.* **41**: 937-946.

40. Togna AP, Shuler ML, Wilson DB. 1993. Effects of plasmid copy number and runaway plasmid replication on overproduction and excretion of beta-lactamase from *Escherichia coli*. *Biotechnol. Prog.* **9**: 31-39.

41. Economou A. 1999. Following the leader: bacterial protein export through the Sec pathway. *Trends Microbiol.* **7**: 315-320.

42. Neumann-Haefelin C, Schafer U, Muller M, Koch HG. 2000. SRP-dependent co-translational targeting and SecA dependent translocation analyzed as individual steps in the export of a bacterial protein. *EMBO J* **19**: 6419-6426.

43. Tian H, Boyd D, Beckwith J. 2000. A mutant hunt for defects in membrane protein assembly yields mutations affecting the bacterial signal recognition particle and Sec machinery. *Proc Natl Acad Sci USA* **97**: 4730-4735.

44. Lee HC, Bernstein HD. 2001. The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proc Natl Acad Sci USA* **98**: 3471– 3476.

45. Peterson JH, Woolhead CA, Bernstein HD. 2003. Basic amino acids in a distinct subset of signal peptides promote interaction with the signal recognition particle. *J Biol Chem.* **278**: 46155-46162.

46. Sijbrandi R. *et al.* 2003. Signal recognition particle (SRP)-mediated targeting and Sec-dependent translocation of an extracellular *Escherichia coli* protein. *J Biol Chem.* **278**: 4654-4659.

47. Patzelt H. *et al.* 2001. Binding specificity of *Escherichia coli* trigger factor. *Proc Natl Acad Sci USA* **98**: 14244-14249.

48. Beck K, Wu LF, Brunner J, Muller M. 2000. Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. *EMBO J* **19**: 134-143.

49. Herskovits AA, Bochkareva ES, Bibi E. 2000. New prospects in studying the bacterial signal recognition particle pathway. *Mol Microbiol*. **38**: 927-939.

50. Fekkes P, Driessen AJ. 1999. Protein targeting to the bacterial cytoplasmic membrane. *Microbiol Mol Biol Rev.* **63**: 161– 173.

51. Gu S-Q, Peske F, Wieden H-J, Rodnina MV, Wintermeyer W. 2003.The signal recognition particle binds to protein L23 at the peptide exit of the *Escherichia coli* ribosome. *RNA* **9**: 566-573.

52. Nagai K, Oubridge C, Kuglstatter A, Menichelli E, Isel C, Jovine L. 2003. Structure, function and evolution of the signal recognition particle. *EMBO J.* **22**: 3479-3485.

53. Koch HG. *et al.* 2001. In vitro studies with purified components reveal signal recognition particle (SRP) and SecA/SecB as constituents of two independent protein-targeting pathways of *Escherichia coli*. *Mol Biol Cell* **10**: 2163–2173.

54. Cristobal S, Scotti P, Luirink J, von Heijne G, de Gier JW. 1999. The signal recognition particle-targeting pathway does not necessarily deliver proteins to the sec-translocase in *Escherichia coli*. *J Biol Chem* **274**: 20068-20070.

55. Scotti PA. *et al.* 1999. SecA is not required for signal recognition particle-mediated targeting and initial membrane insertion of a nascent inner membrane protein. *J Biol Chem.* **274**: 29883–29888.

56. Qi HY, Bernstein HD. 1999. SecA is required for the insertion of inner membrane proteins targeted by the *Escherichia coli* signal recognition particle. *J Biol Chem*. **274**: 8993-8997.

57. Bowers CW, Lau F, Silhavy TJ. 2003. Secretion of LamB–LacZ by the signal recognition particle pathway of *Escherichia coli*. *J Bacteriol.* **185**: 5697-5705.

58. Lee HC, Bernstein HD. 2001. The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proc Natl Acad Sci USA* **98**: 3471-3476.

59. Stanley NR, Palmer T, Berks BC. 2000. The twin arginine consensus motif of Tat signal peptides is involved in Sec independent protein targeting in *Escherichia coli*. *J Biol Chem*. **275**: 11591-11596.

60. Yahr TL,Wickner WT. 2001. Functional reconstitution of bacterial Tat translocation in vitro. *EMBO J.* **20**: 2472–2479.

61. de Leeuw E. *et al.* 2002. Oligomeric properties and signal peptide binding by *Escherichia coli* Tat protein transport complexes. *J Mol Biol.* **322**: 1135-1146.

62. Bogsch EG, Sargent F, Stanley NR, Berks BC, Robinson C, Palmer T. 1998. An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. *J Biol Chem.* **273**: 18003-18006.

63. Santini CL. *et al.* 2001. Translocation of jellyfish green fluorescent protein via the Tat system of *Escherichia coli* and change of its periplasmic localization in response to osmotic up-shock. *J Biol Chem.* **276**: 8159-8164.

64. Sargent F, Stanley NR, Berks BC, Palmer T. 1999. Sec-independent protein translocation in *Escherichia coli* A distinct and pivotal role for the TatB protein. *J Biol Chem.* **274**: 36073-36082.

65. De Lisa MP, Tullman D, Georgiou G. 2003. Folding quality control in the export of proteins by the bacterial twin arginine translocation pathway. *Proc Natl Acad Sci USA*. **100**:6115-6120.

66. Barrett CM, Ray N, Thomas JD, Robinson C, Bolhuis A. 2003. Quantitative export of a reporter protein, GFP, by the twin-arginine translocation pathway in *Escherichia coli*. *Biochem Biophys Res Commun.* **304**: 279-284.

67. De Lisa MP, Samuelson P, Palmer T, Georgiou G. 2002. Genetic analysis of the twin arginine translocator secretion pathway in bacteria. *J Biol Chem.***277**: 29825-29831.

68. Palmer T, Berks BC. 2003. Moving folded proteins across the bacterial cell membrane. *Microbiology*. **149**: 547-556.

69. Sargent F. *et al.* Purified components of the *Escherichia coli* Tat protein transport system form a double-layered ring structure. *Eur J Biochem* **268**: 3361-3367.

70. Allen SC, Barrett CM, Ray N, Robinson C. 2002. Essential cytoplasmic domains in the *Escherichia coli* TatC protein. *J Biol Chem.* **277**: 10362-10366.

71. de Leeuw E. *et al.* 2002. Oligomeric properties and signal peptide binding by *Escherichia coli* Tat protein transport complexes. *J Mol Biol.* **322**: 1135-1146.

72. Schnell DJ, Hebert DN. 2003. Protein translocons: multifunctional mediators of protein translocation across membranes. *Cell* **112**: 491-505.

73. Alami M. *et al.* 2003. Differential interactions between a twin arginine signal peptide and its translocase in *Escherichia coli*. *Mol Cell.* **12**: 937-946.

74. Palmer T, Sargent F, Berks BC. 2004. Light traffic: photo-crosslinking a novel transport system. *Trends Biochem Sci.* **29**: 55-57.

75. Berks BC, Sargent F and Palmer T. 2000. The Tat protein export pathway. *Molecular Microbiology* **35**: 260-274.

76. Blaudeck N, Sprenger GA, Freudl R, Wiegert T. 2001. Specificity of signal peptide recognition in tat-dependent bacterial protein translocation. *J Bacteriol*. **183**: 604-610.

77. Ize B. *et al.* 2002. In vivo dissection of the Tat translocation pathway in *Escherichia coli*. *J Mol Biol.* **317**: 327-335.

78. Oresnik IJ, Ladner CL, Turner RJ. 2001. Identification of a twin-arginine leader-binding protein. *Mol Microbiol.* **40**: 323-331.

79. Robinson C, Bolhuis A. 2004. Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochim. Biophys. Acta.* **1694**: 135-147.

80. Rose RW, Brüser T, Kissinger JC, Pohlschröder M. 2002. Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol. Microbiol.* **45**: 943–950.

81. de Keyzer J, van der Does C, Driessen AJM. 2003. The bacterial translocase: a dynamic protein channel complex, *Cell. Mol. Life Sci.* **60**: 2034-2052.

82. Berks BC, Palmer T, Sargent F. 2005. Protein targeting by the bacterial twinarginine translocation (Tat) pathway, *Curr. Opin. Microbiol.* **8**: 174-181.

83. Jack RL, Buchanan G, Dubini A, Hatzixanthis K, Palmer T, Sargent F. 2004. Coordinating assembly and export of complex bacterial proteins, *EMBO J.* **23**: 3962-3972.

84. Sanders C, Wethkamp N, Lill H. 2001. Transport of cytochrome c derivatives by the bacterial Tat protein translocation system, *Mol. Microbiol.* **41**: 241-246.

85. Lory S. 1998. Secretion of proteins and assembly of bacterial surface organelles: shared pathways of extracellular protein targeting. *Curr Opin Microbiol.* **1**: 27-35.

86. Pugsley AP, Francetic O, Possot OM, Sauvonnet N, Hardie K. 1997. Recent progress and future directions in studies of the main terminal branch of the general secretory pathway in Gram-negative bacteria—a review. *Gene* **192**: 13-19.

87. Sandkvist M. Biology of type II secretion. 2001. *Mol Microbiol*. **40**: 271–83.

88. Nouwen N. *et al.* 1999. Secretin PulD: association with pilot PulS, structure, and ion-conducting channel formation. *Proc Natl Acad Sci USA* **96**: 8173-8177.

89. Possot OM, Vignon G, Bomchil N, Ebel F, Pugsley AP. 2000. Multiple interactions between pullulanase secreton components involved in stabilization and cytoplasmic membrane association of PulE. *J Bacteriol.* **182**: 2142-2152.

CHAPTER 3

GENOMICS AND PROTEOMICS IN PROCESS DEVELOPMENT: OPPORTUNITIES AND CHALLENGES

3.1 Preface

There have been increasing efforts to apply genome-wide expression profiling tools to understand cell culture process development with the goal of strain improvement or process improvement. This chapter is adapted from: Gupta, P., and Lee, K.H. 2007. Genomics and Proteomics in Process Development: Opportunities and Challenges. *Trends in Biotechnology* 25: 324-330 and presents a review of the application of genomics and proteomics in process development.

3.2 Abstract

Global gene expression profiling by genomic and proteomic analyses has changed the face of drug discovery and biological research in the past few years. The impact of these technologies in the area of process development for recombinant protein production has been increasingly realized. This review discusses the application of genome-wide expression profiling tools in the design and optimization of bioprocesses with an emphasis on the impact on process development of mammalian cell culture. Despite a lack of genome sequence information for most relevant mammalian cell lines used, these technologies can be applied during various process development steps. Although, there are only a few examples in the literature that present a significant improvement in productivity based on genomics and proteomics, further advances in analytical tools and genome sequencing technologies will greatly increase our knowledge at the molecular level and will drive the design of future bioprocesses.

3.3 Introduction

The market for biopharmaceuticals is estimated to increase from USD$33 billion to more than USD$70 billion by the end of the decade [1]. Overall, some 165 biopharmaceutical products (recombinant proteins, monoclonal antibodies and nucleic-acid based drugs) have Food and Drug Administration (FDA) approval and several thousand are in discovery and pre-clinical development [1]. As a result there is a need for efficient, cost-effective systems for the production of these molecules.

Most biopharmaceuticals are produced through the use of recombinant DNA technology, wherein a recombinant 'production system' is created based on a genetically modified host cell. These production systems usually involve either microbial fermentation or mammalian cell culture among other platform technologies. In addition to fermentation, the extraction and subsequent purification of desired proteins from the fermentation broth is an important part of the overall production process of biopharmaceuticals. These downstream procedures form an integral part of the overall process development (Fig. 3.1) and are crucial in determining the final productivity and characteristics of the product. It is essential that a suitable production process is designed before pre-clinical trials and that the process is scalable and yields a sufficient amount of therapeutic protein. Hence, extensive early development work is essential. Traditionally, process development involves designing and optimizing upstream and downstream processes based on empirical data, with only incomplete or limited understanding at the cellular level. The advent of genomic and proteomic technologies, which can provide molecular level details, has generated interest in the application of these tools to process development. This review discusses the role of these approaches in the context of improving different stages of process development, including both upstream and downstream processes. The applications to mammalian

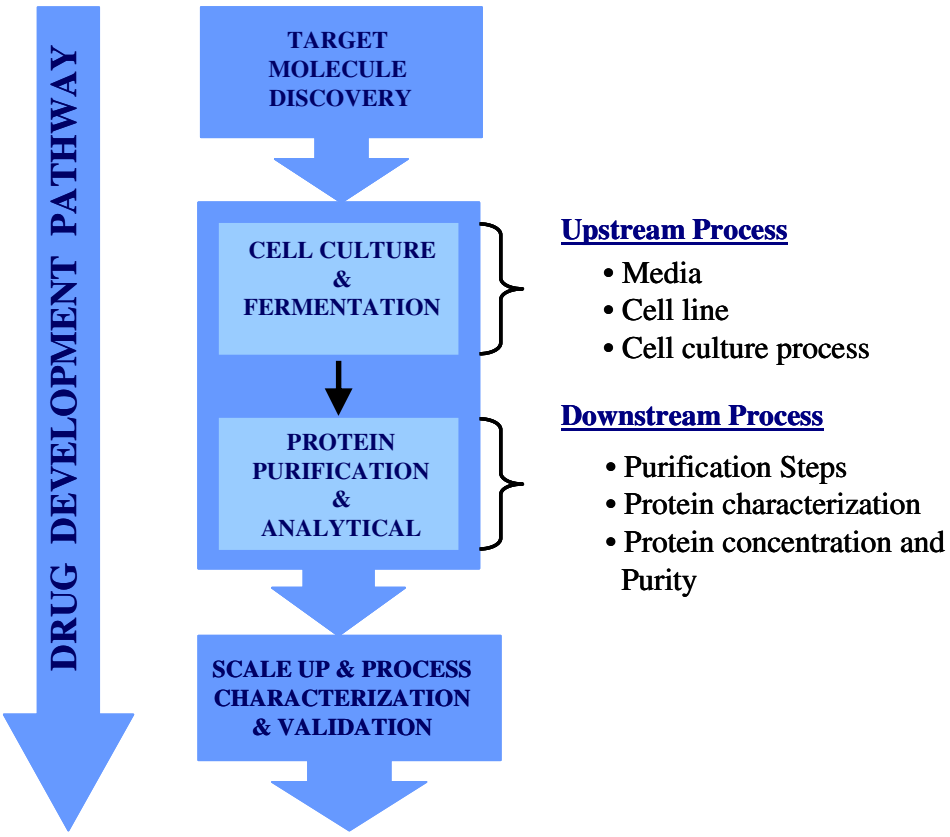


Figure 3.1. Development pathway for therapeutic proteins.

cell culture are emphasized because production processes for these cells are more complex than for microbial culture and because there is a growing importance on the use of mammalian cells as a production platform.

3.4 Role of Genomics and Proteomics

Genomics is the comprehensive analysis of the genetic content of an organism and also often refers to genome-wide studies of mRNA expression [2] Proteomics is the study of the expressed protein complement of a genome at a specific time [3]. The drivers of genomic and proteomic analyses are the technological achievements of the past decade that enable a reasonably quantitative analysis of DNA sequence, mRNA, and protein expression inside cells and include tools such as DNA microarrays, two-dimensional gel electrophoresis (2DE), and mass spectrometry (MS). The application of these methods in drug discovery has heralded a new era for target identification and these efforts have been reviewed recently [4-7]. These genome-wide approaches have been adopted by biotechnology and pharmaceutical industries to complement traditional approaches to target identification and validation, for hypotheses generation, and for experimental analyses in traditional-based methods.

An equally important, but less-often discussed, issue is the impact of genomics and proteomics on process development (Fig. 3.2). From the process development perspective, these tools might significantly aid in at least three areas: 1) exploration of cellular functions to enhance productivity or influence desired properties of biological products (**upstream**), 2) learn and apply knowledge of cell function in response to environmental change, including exposure to normal and unusual substrates, and 3) exploit knowledge of cell function and properties to improve product purification and characterization (**downstream**).

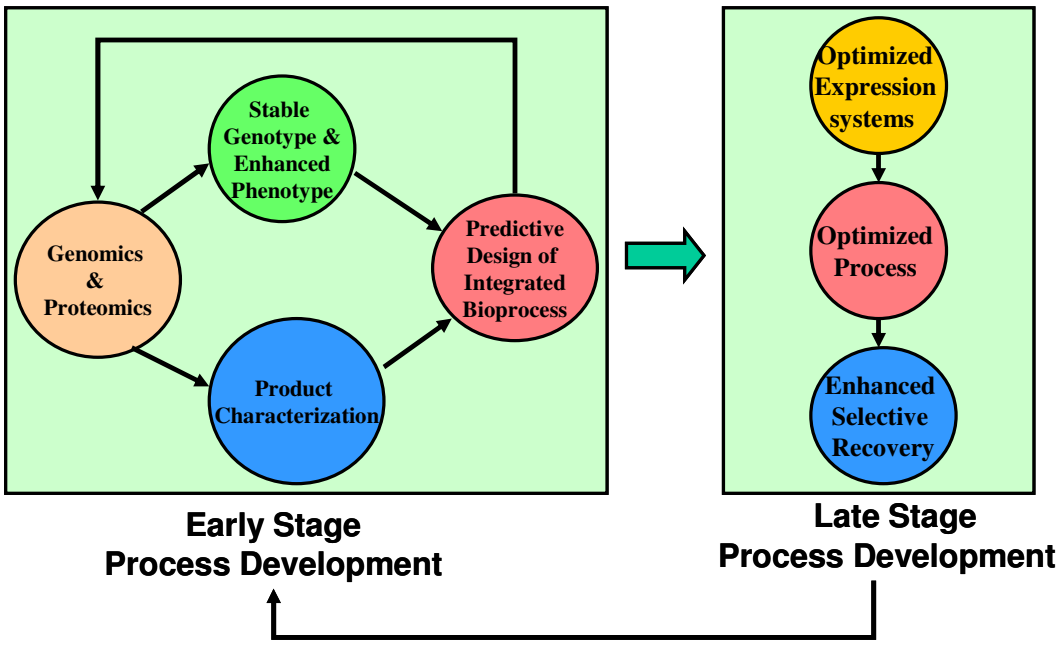


Figure 3.2. A possible role of genomic and proteomic tools in process development. Fundamental understanding of genomic and proteomic processes can facilitate the selection of productive organisms. It can also facilitate product/impurity characterization between different cell lines, leading to better integration of upstream and downstream processing operations. In theory, further improvements may be made using genomic and proteomic analyses, in an iterative fashion. These resulting systems can then be used by late stage process development to validate and scale up the process.

3.5 Mammalian Cell Culture

The advances made in mammalian cell culture technology during the past two decades have been greatly facilitated by genetic and physiological manipulation of cells. Today, approximately 60-70% of all recombinant protein pharmaceuticals are produced in mammalian cells [8]. The major mammalian cell lines that have gained regulatory approval for recombinant protein production are Chinese hamster ovary (CHO), mouse myeloma (NSO), baby hamster kidney (BHK), human embryo kidney (HEK-293) and human retinal cells [8]. Of these, CHO is the most widely used cell line. However, the genetic and physiological properties that enable CHO cells to be capable producers are not fully understood. Although whole genome microarrays are commercially available for mouse and human cell lines, there is a lack of genome sequence information available for CHO [9]. Progress has been made in the large-scale expressed sequence tag (EST) sequencing of cultured mammalian cell lines designed specifically for recombinant protein production [10]. In that effort, two cDNA libraries were constructed from three CHO cell lines grown under different media conditions. These libraries led to the construction of a cDNA microarray containing 4608 ESTs, which yielded 2602 unique assemblies upon sequencing, of which 76% were annotated as orthologs of sequences in the GenBank database (http://www.ncbi.nlm.nih.gov/Genbank). This initiative also revealed that CHO sequences are generally most similar to those of the mouse. It is very likely that with advances in sequencing technologies, we will see more refined and accurate cDNA microarrays for this important cell line in the near future.

3.5.1 Upstream Process Development

Cell Culture Media

A significant challenge in process development is to establish well-defined

manufacturing processes that are robust and reproducible. One of the most important factors that can affect a cell culture process is the culture media. Fetal Bovine Serum (FBS) has been the most widely used growth supplements for cell cultures, primarily because of its high levels of growth stimulatory factors. However, the quality, type, and concentration of the components in different FBS lots can affect the cellular growth rate in a fermentation process [11]. In a recent study, the growth rates of adult retinal pigment epithelial cells (ARPE-19) were evaluated using three different lots of FBS [12]. The authors used proteomic techniques, based on reverse-phase liquid chromatography (LC) coupled with ion-trap tandem MS analysis, to examine the variability of important growth stimulatory and inhibitory proteins in different FBS serum lots. The study revealed that serum lots resulting in the highest growth rates contained additional growth factors and related proteins that were not found in the other two serum lots.

Currently, a major focus in cell culture media development is the formulation of serum-free and animal-product-free media that can result in consistent growth and productivity without disease transmission. High-throughput genomic and proteomic tools can have a vital role in medium development and optimization. A recent study, using microarray analysis, identified specific receptors, cell adhesion molecules, and cell-signaling factors expressed during cell growth [13]. This information led to the identification of corresponding ligands and small molecules that might be incorporated into media to test for the desired effect on given cellular processes. Another recent study involving the development of protein-free media suggested that zinc metal could be used as an insulin replacement in murine hybridoma cultures [14]. In this transcript analysis, using mouse oligonucleotide array and quantitative real time reverse transcriptase polymerase chain reaction (RT-PCR), indicated no major change in the

global expression profile between the insulin and zinc supplemented cultures, which is consistent with their similar growth and metabolic characteristics and monoclonal antibody production profiles.

Cell line selection and engineering

The ability to produce and select for a high producing cell line is key to the initial stages of bioprocess development. In the case of mammalian cells, the recombinant gene with the necessary transcription regulatory elements is transferred to cells along with a second gene that confers a selective advantage to the recipient cells. Most commonly, dihydrofolate reductase and glutamine synthetase genes are used for selection in CHO cells. To improve recombinant protein production from mammalian cells, various molecular strategies have been used at the level of gene copy number, transcription, translation, posttranslational modification and secretion, as reviewed by [8]. Despite these studies, the bottlenecks in the cellular machinery for efficient recombinant protein production are still unclear.

The site of incorporation of the recombinant genes within the genome of the host mammalian cell is crucial in determining gene stability and productivity [15-16]. Techniques for the introduction of recombinant genes into mammalian cells is a random process and results in a clone-to-clone variation in protein productivity for a given cell line, resulting in a wide range (may exceed two orders of magnitude) of protein expression levels of different cell clones [17-18]. Recently, a study analyzed how GS-NS0 mammalian cell transfection produces variant cell lines with distinct characteristics in terms of productivity of monoclonal antibody anti-CD38 [19]. The authors used genomic quantitation approaches, Northern and Southern analysis, to define molecular rationales for high and low level producers and to describe processes

that allow cells to escape the stringency of selection procedures. It was found that >50% of the transfectants studied had molecular defects at the level of cDNA and/or mRNA, including defects in the regulatory regions. A similar study explored the regulation of recombinant monoclonal humanized IgG production in CHO cells by a comparative study of gene copy number, mRNA level, and protein expression between two different cell families, each containing one parental cell line and two progeny cell lines (amplified in the presence of methotrexate) [20]. The authors observed that progeny cell lines in both families had higher productivities than parental cell lines and concluded that it might be as a result of gene amplification and enhanced transcriptional efficiency.

A proteomics study of whole cell extracts from four stably transfected GS-NS0 cell lines that differ in recombinant monoclonal antibody IgG_4 productivity (qMAb) used 2DE coupled with MS [21]. A significant increase in abundance of proteins including molecular chaperone (BiP), endoplasmin, and protein disulphide isomerase was measured with increasing monoclonal antibody (MAb) productivities. In another study, the functional categorization of 76 proteins, which were differentially expressed among the same four GS-NS0 cell lines, was also presented [22]. The authors revealed that protein synthesis, degradation, and nucleic acid synthesis and processing protein categories did not change in abundance. However, ER chaperones, non ER-chaperones, and cytoskeletal protein categories all increased significantly with elevated qMAb, inferring that the rate of production of MAb in GS0 cells is limited by the availability of processing and/or secretory apparatus of the cell.

Genomic and proteomic approaches can also offer new insights into cellular metabolism and physiology for metabolic engineering. This type of combined

approach has been used to understand the molecular mechanism of metabolic shift in the mouse hybridoma cell line (MAK) [23]. Metabolic shift is a consequence of controlled nutrient feeding to maintain low concentrations of glucose and glutamine, which results in higher cell concentration in continuous cultures. In this study, two cell lines with different ratios of glucose consumption to lactate production were studied using mouse cDNA microarrays to identify differentially expressed mRNA transcripts and 2DE-MS to identify differentially expressed proteins. It was suggested that metabolic shift is a combined effect of both biochemical events at the metabolic reaction level and gene expression at the transcription and translation levels. This approach of integrating transcriptional profiling, proteomic techniques and biochemical analysis constitutes what has been termed as systems biology [24-25] and can provide a comprehensive understanding of mammalian cells in culture.

Recently, large-scale gene expression analysis was performed to better understand the cholesterol-dependent phenotype of NS0 myeloma cells [26]. Transcriptional analysis between a cholesterol-dependent cell line and a cholesterol-independent cell line was performed using mouse Affymetrix GeneChips^{TM}. Proteomic analysis was performed using 2DE coupled with MS. Most of the genes involved in cholesterol biosynthesis, lipid metabolism, and central energy metabolism were expressed at lower levels in cholesterol-independent cell line, indicating that the reversal of cholesterol dependency has a profound effect on cell physiology.

2DE has also been used to analyze the simultaneous expression of several important cell cycle regulatory proteins [27]. This study found that cyclin D1, cyclin E, and E2F-1 proteins appear to have the strongest correlation to the mitogenic strength of growth stimulation of CHO cells. The authors hypothesized that the deregulated expression of

these proteins might bypass the requirement for external growth factor signaling. Transfection with a vector encoding cloned cyclin E and overexpression of cloned E2F-1 have independently been shown to activate proliferation of CHO cells in the absence of serum and external growth factors [28-29]. The new insights gained from all of these studies will continue to facilitate the development of advanced host strains, new processes, and new strategies of process development.

Cell Culture Conditions

Temperature

One area in which there is significant interest is the use of reduced temperature cultivation for enhanced protein production in mammalian cells. The effect of culture temperature on CHO cell growth and recombinant protein productivity has been investigated extensively [30-34]. These studies revealed that low culture temperature, in a range of 28-34ºC, improves protein productivity, even though specific growth rate is decreased. To exploit these responses fully, the molecular responses governing cellular adaptation to cold-shock in mammalian cells needs to be better understood.

Kaufmann and co-workers used 2DE to show that CHO cells respond to low culture temperature (30ºC) by synthesizing specific cold-inducible proteins, which might arrest cell proliferation in the G1 phase of the cell cycle [35]. Recently, transcriptome and proteome analyses of low temperature-induced expression in CHO cells, producing erythropoietin, were performed [36]. Proteomic analysis identified nine proteins, including some chaperones, with differential expression under conditions of low-temperature (33ºC). Genomic analysis using rat and mouse cDNA arrays revealed differential gene expression of various cellular processes including metabolism, transport, signaling etc. However, of the nine proteins identified by proteomic

analysis, the mRNA transcript expression patterns of only four proteins were detected in rat arrays. Similarly only two proteins were detected in the mouse arrays.

Hyperosmotic Pressure

Genomics and proteomics can also be used to help understand intracellular and physiological changes in cells and to obtain better insight into possible environmental or genetic manipulations for increasing productivity. Environmental manipulation of hyperosmotic pressure, which can be induced by adding salts or sugars to culture media, is one strategy thought to be highly feasible for improving desired protein productivity in recombinant CHO cell culture [37-38]. However, because cell growth is compensated at elevated osmolarity, the specific protein productivity does not increase substantially. To better understand the intracellular response of CHO cells to hyperosmotic pressure, proteomics of CHO cells producing a chimeric antibody under hyperosmotic pressure was performed [39]. Interestingly, under hyperosmotic conditions in which sodium chloride was added to a final osmolality of 450 mOsm/kg-media, the authors found an increase in protein expression of pyruvate kinase and glyceraldehydes-3-phosphate dehydrogenase and decrease in tubulin expression. Separately, a genome-wide analysis of the transcriptional response of the murine hybridoma cell line, OKT3, towards hyperosmotic stress using DNA microarrays was investigated [40]. The authors obtained a list of 215 genes that were differentially expressed in response to osmotic shock, including genes related to immunoglobins, metabolism/catabolism, cell cycle regulation, signaling pathways, etc.

Small chemical compounds

The effect of small chemical molecules on recombinant protein productivity and quality is also an important area of research and has benefited from genomic and

proteomics approaches. Cytochalasin D, a fungal metabolite, enhanced protein productivity in CHO cells. In a study of CHO cells, a proteomic comparison was made between a productive recombinant clone selected at high methotrexate (MTX) concentration and clones of intermediate and low productivities [41]. The authors used this data and noted a fourfold increase in actin-capping protein (CapZ). Because the function of CapZ is similar to the effects of a small molecule cytochalasin D, the authors hypothesized that the addition of cytochalasin D might result in enhanced productivity and product secretion. In combination with MTX gene amplification, the addition of cytochalasin D resulted in a 52- to 150-fold increase in recombinant protein productivity.

A proteomics approach was used to investigate the effects of butyrate and zinc sulphate on the overall proteome of a recombinant CHO cell line producing human growth hormone were investigated using a proteomics approach [42]. This study found that the addition of these compounds induced metabolic and stress protection proteins. Another recent study evaluated the effect of elevated ammonium ion concentration (media supplemented with 10mM ammonium choloride) on 12 glycosylation related genes in CHO cells, producing constitutively expressed tissue plasminogen activator (t-PA) [43]. The authors, using quantitative real time RT-PCR, observed decreased gene expression for cytosine monophosphate-sialic acid transporter, β(1,4)-galactosyltransferase, uridine diphosphate-glucose pyrophosphorylase and α(2,3)-sialyltransferase proteins for the ammonium treated culture. The study concluded that lower levels of these key glycosylation enzymes and transporters cause a higher molecular heterogeneity of t-PA glycoforms under ammonium stress.

3.5.2 Downstream Process Development

The application of –omics technologies in protein purification and characterization has been an active area of interest. However, as compared to the cell culture end of the process, the impact of these technologies on the downstream end of the process development has not been fully realized. The lack of significant studies, impacting this important process, might be a result of the biochemical nature (versus cellular nature) of these processes.

The paradigm of routine biomanufacturing is to ensure batch-to-batch consistency in product purity, efficacy, and safety. Alterations to structural features such as protein fold or post-translational modification (glycosylation, phosphorylation, etc.) can alter product efficacy. It is thus important to identify and monitor these critical parameters during production to assess the effect of changes in cell culture conditions on product quality. In particular, the number of oligosaccharide structures associated with specific fermentation can be highly variable [44-45]. Traditionally, protein gel electrophoresis has been used to assess purity and integrity of the secreted product isolated from the culture supernatant [46]; LC and MS have been used to identify the protein product and to study its structure and glycosylation [47-48]. An important advancement in the characterization of glycosylation is the use of capillary electrophoresis-MS. This technique was used recently to accurately determine 44 glycoforms of recombinant human erythropoietin with high mass accuracy (± 1 Dalton) [49]. Despite the precision of these technologies, all of these approaches depend on extensive purification protocols, which might be time consuming and expensive.

Recently, a proteomic technique was used to characterize overexpressed recombinant t-PA from a CHO cell culture lysate [50]. The authors used reversed phase capillary

LC-linear ion trap/Fourier transform MS to identify and characterize all the major glycoforms of t-PA [51], with 92.2% sequence coverage. This technology can be helpful in monitoring the glycosylation patterns during the fermentation process, thereby alleviating the need of purification for product characterization. Industrially, this strategy can be very useful in the development of process analytical technology applications, in which glycosylation can be a key indicator of cell viability as well as product quality.

The removal of CHO-derived proteins from the biopharmaceutical product is often monitored using multi-product immunoassays [52-53]. To address whether underlying differences between CHO cell lines result in sufficient protein expression changes, a comparative proteomics study of three independently generated CHO cell lines was performed [54]. This study yielded only 11 qualitative changes and 26 quantitative changes greater than two-fold in protein expression out of a total of 1000 statistically analyzed proteins. Identification of these protein spots by MS revealed that many of the observed changes were due to post-translational modifications. These results support the idea of using multi-product immunoassays to monitor host cell protein impurities in different CHO cell lines.

3.6 Microbial Cell Culture

As could be expected, the literature related to industrially relevant microorganisms is greater than that for mammalian cells lines. We will refer the reader to several excellent papers and reviews, which illustrate the effect of –omics technologies in the process development of microbial cell culture. Genome sequencing and annotation of many industrially relevant microorganisms forms the foundation for relatively simple strain and process development. Expression profiling such microorganisms generates

valuable information that can be used for the development of metabolic and cellular engineering strategies to enhance the yield and productivity of recombinant proteins and to modify cellular properties to impact process development, as reviewed previously [55-57].

There have been several reports on proteome profiling [58-66] and transcriptome profiling [67-72] of *Escherichia coli* strains producing recombinant proteins. Apart from investigating the effects of culture conditions (media substrates, temperature, pH etc.), many of these studies present a metabolic rationale to increase the recombinant protein production, often by co-expression of one or more proteins of specified activity along with the protein of interest. Of particular interest is a transcriptomic and proteomic analysis of global physiological changes in *E. coli* during high cell density cultivation (HCDC) [73]. In this study, DNA microarray and 2DE analyses revealed that the expression of amino acid biosynthesis genes was decreased as cell density increases, which might be the cause for reduced productivity of recombinant proteins during HCDC. However, the expression of chaperones was increased during HCDC, suggesting that the high-density condition can be stressful to the cells. This type of global analysis can provide invaluable information in developing metabolic engineering and fermentation strategies in industrially relevant HCDC.

3.7 Future Perspectives

New proteomics and genomics technologies such as shotgun proteomics and other quantitative methods have been used extensively in drug discovery and understanding the complex cellular processes, as reviewed in [74-75]. Although, these emerging methods have not yet been applied comprehensively in the process development because they are relatively new, these technologies have the potential to increase the

impact of genomics and proteomics on the design of future bioprocesses.

3.8 Conclusion

Advances in genomic and proteomic technologies are transforming the drug development process. While there are challenges associated with the collection and analysis of large datasets generated by such technological applications, a topic not discussed in this review, genome-wide analysis of expression profiles can be greatly beneficial in understanding living systems. These beneficial effects on the industry are more clear at the discovery level because of the relatively larger investment made at that level. However, the impact on process development and manufacturing should not be underestimated. There are increasing efforts to apply these technologies to understanding cell culture process development with the goal of strain improvement or process improvement. Unfortunately, a significant number of these efforts result only in "lists of proteins" that change significantly and fall short of demonstrating the impact of the controlled expression of those genes or proteins on desired phenotypes. Thus, although there is considerable work being done in using these methods and there is great promise in the application of these methods, there is *at this time*, relatively few examples that demonstrate the significant impact of genomics and proteomics on process development.

3.9 Acknowledgements

We thank Dana Andersen for insightful comments. This work was supported in part by the New York State office of Science, Technology, and Academic Research and by Cornell University.

REFERENCES

1. Walsh G. 2006. Biopharmaceutical benchmarks 2006. *Nat Biotechnol.* **24**: 769-76.

2. Lee PS, Lee KH. 2000. Genomic Analysis. *Curr Opin Biotechnol.* **11**: 171-175.

3. Dutt MJ, Lee KH. 2000. Proteomic Analysis. *Curr Opin Biotechnol.* **11**: 176-179.

4. Burbaum J, Tobal GM. 2002. Proteomics in drug discovery. *Curr Opin Chem Biol.* **6**: 427-33.

5. Butcher EC, Berg EL, Kunkel EJ. 2003. Systems biology in drug discovery. *Nat Biotechnol.* **22**: 1253-9.

6. Kramer R, Cohen D. 2004. Functional genomics to new drug targets. *Nat Rev Drug iscov*. **3**: 965-72.

7. Onyango P. 2004. The Role of Emerging Genomics and Proteomics Technologies in Cancer Drug Target Discovery. *Current Cancer Drug Targets*. **4**: 111-124.

8. Wurm FM. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol*. **22**: 1393-8.

9. Wlaschin KF, Seth G, Hu WS. 2006. Toward genomic cell culture engineering. *Cytotechnology*. **50**: 121-140.

10. Wlaschin KF, Nissom PM, Gatti Mde L, Ong PF, Arleen S, Tan KS, Rink A, Cham B, Wong K, Yap M, Hu WS. 2005. EST sequencing for gene discovery in Chinese hamster ovary cells. *Biotechnol Bioeng*. **91**: 592-606.

11. Boone CW, Mantel N, Caruso TD Jr, Kazam E, Stevenson RE. 1971. Quality control studies on fetal bovine serum used in tissue culture. *In Vitro*.**7**: 174-89.

12. Zheng X, Baker H, Hancock WS, Fawaz F, McCaman M, Pungor E Jr. 2006. Proteomic Analysis for the Assessment of Different Lots of Fetal Bovine Serum as a Raw Material for Cell Culture. Part IV. Application of Proteomics to the Manufacture of Biological Drugs. *Biotechnol Prog*. **22**: 1294-300.

13. Allison DW, Aboytes KA, Fong DK, Leugers SL, Johnson TK, Loke HN, Donahue LM. 2005. Development and optimization of cell culture media—genomic and proteomic approaches. *BioProcess Int.* **3**:2–7.

14. Wong VV, Nissom PM, Sim SL, Yeo JH, Chuah SH, Yap MG. 2006. Zinc as an insulin replacement in hybridoma cultures. *Biotechnol Bioeng.* **93**: 553-63.

15. Wurm FM, Petropoulos CJ. 1994. Plasmid integration, amplification and cytogenetics in CHO cells: questions and comments. *Biologicals.* **22**: 95-102.

16. Yoshikawa T, Nakanishi F, Ogura Y, Oi D, Omasa T, Katakura Y, Kishimoto M, Suga K. 2000. Amplified gene location in chromosomal DNA affected recombinant protein production and stability of amplified genes. *Biotechnol Prog.* **16**: 710-5.

17. Barnes LM, Bentley CM, Dickson AJ. 2004. Molecular definition of predictive indicators of stable protein expression in recombinant NS0 myeloma cells. *Biotechnol Bioeng.* **85:** 115−121

18. Jones D, Kroos N, Anema R, van Montfort B, Vooys A, van der Kraats S, van der Helm E, Smits S, Schouten J, Brouwer K, Lagerwerf F, van Berkel P, Opstelten DJ, Logtenberg T, Bout A. 2003. High level expression of recombinant IgG in the human cell line PER.C6. *Biotechnol Prog.* **19**: 163−168.

19. Barnes LM, Bentley CM, Moy N, Dickson AJ. 2007. Molecular analysis of successful cell line selection in transfected GS-NS0 myeloma cells. *Biotechnol Bioeng*. **96**: 337-48.

20. Jiang Z, Huang Y, Sharfstein ST. 2006. Regulation of recombinant monoclonal antibody production in Chinese hamster ovary cells: A comparative study of gene copy number, mRNA level, and protein expression. *Biotechnol. Prog.* **22**: 313-318.

21. Smales CM, Dinnis DM, Stansfield SH, Alete D, Sage EA, Birch JR, Racher AJ, Marshall CT, James DC. 2004. Comparative proteomic analysis of GS-NS0 murine myeloma cell lines with varying recombinant monoclonal antibody production rate. *Biotechnol Bioeng.* **88**: 474-88.

22. Dinnis DM, Stansfield SH, Schlatter S, Smales CM, Alete D, Birch JR, Racher AJ, Marshall CT, Nielsen LK, James DC. 2006. Functional proteomic analysis of GS-NS0 murine myeloma cell lines with varying recombinant monoclonal antibody production rate. *Biotechnol Bioeng.* **94**: 830-41.

23. Korke R, Gatti Mde L, Lau AL, Lim JW, Seow TK, Chung MC, Hu WS. 2004. Large scale gene expression profiling of metabolic shift of mammalian cells in culture. *J Biotechnol.* **107**: 1-17.

24. Ideker T, Galitski T, Hood L. 2001. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet.* **2**:343-72.

25. Kuczenski R, Aggarwal K, Lee KH. 2005. Improved Understanding of Gene Expression Regulation Using Systems Biology. *Expert Rev Proteomics.* **2**: 915-924.

26. Seth G, Philp RJ, Denoya CD, McGrath K, Stutzman-Engwall KJ, Yap M, Hu WS. 2005. Large-scale gene expression analysis of cholesterol dependence in NS0 cells. *Biotechnol Bioeng.* **90**: 552-67.

27. Lee KH, Harrington MG, Bailey JE. 1996b. Two-Dimensional Electrophoresis of Proteins as a Tool in the Metabolic Engineering of Cell Cycle Regulation. *Biotechnol Bioeng*. **50**: 336-340.

28. Lee KH, Sburlati A, Renner WA, Bailey JE. 1996a. Deregulated expression of cloned transcription factor E2F-1 in Chinese hamster ovary cells shifts protein patterns and activates growth in protein-free medium. *Biotechnol Bioeng*. **50**: 273-79.

29. Renner WA, Lee KH, Hatzimanikatis V, Bailey JE, Eppenberger HM. 1995. Recombinant Cyclin E Expression Activates Proliferation and Obviates Surface Attachment of Chinese Hamster Ovary (CHO) Cells in Protein-Free Medium. *Biotechnol Bioeng*. **47**: 476-82.

30. Bloemkolk JW, Gray MR, Merchant F and Mosmann TR. 1992. Effect of temperature on hybridoma cell cycle and MAb production. *Biotechnol Bioeng.* **40**: 427–431.

31. Fox SR, Patel UA, Yap MG, Wang DI. 2004. Maximizing interferon-gamma production by Chinese hamster ovary cells through temperature shift optimization: experimental and modeling. *Biotechnol Bioeng.* **85**: 177-84.

32. Furukawa K, Ohsuye K. 1998. Effect of culture temperature on a recombinant CHO cell line producing a C-terminal α-amidating enzyme. *Cytotechnology.* **26**: 153-164.

33. Hendrick V, Winnepenninckx P, Abdelkafi C, Vandeputte O, Cherlet M, Marique T, Renemann G, Loa A, Kretzmer G, Werenne J. 2001. Increased productivity of recombinant tissular plasminogen activator (t-PA) by butyrate and shift of temperature: a cell cycle phases analysis. *Cytotechnology*. **36**: 71-83.

34. Al-Fageeh MB, Marchant RJ, Carden MJ, Smales CM. 2006. The cold-shock response in cultured mammalian cells: harnessing the response for the improvement of recombinant protein production. *Biotechnol Bioeng.* **93**: 829-35.

35. Kaufmann H, Mazur X, Fussenegger M, Bailey JE. 1999. Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells. *Biotechnol Bioeng*. **63**: 573-82.

36. Baik JY, Lee MS, An SR, Yoon SK, Joo EJ, Kim YH, Park HW, Lee GM. 2006. Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin. *Biotechnol Bioeng.* **93**: 361-71.

37. Chen Z, Liu H, WU B. 1998. Hyperosmolality leads to an increase in tissue-type plasminogen activator production by a Chinese hamster ovary cell line. *Biotechnol. Tech.* **12**: 207-209.

38. Ryu JS, Kim TK, Chung JY, Lee GM. 2000. Osmoprotective effect of glycine betaine on foreign protein production in hyperosmotic recombinant chinese hamster ovary cell cultures differs among cell lines. *Biotechnol Bioeng.***70**: 167-75.

39. Lee MS, Kim KW, Kim YH, Lee GM. 2003. Proteome analysis of antibody-expressing CHO cells in response to hyperosmotic pressure. *Biotechnol Prog.* **19**: 1734-41.

40. Shen D, Sharfstein ST. 2006. Genome-wide analysis of the transcriptional response of murine hybridomas to osmotic shock. *Biotechnol Bioeng*. **93**: 132-45.

41. Hayduk EJ, Lee KH. 2005. Cytochalasin D can improve heterologous protein productivity in adherent Chinese hamster ovary cells. *Biotechnol Bioeng*. **90**: 354-64.

42. Van Dyk DD, Misztal DR, Wilkins MR, Mackintosh JA, Poljak A, Varnai JC, Teber E, Walsh BJ, Gray PP.2003. Identification of cellular changes associated with increased production of human growth hormone in a recombinant Chinese hamster ovary cell line. *Proteomics*.**3**: 147-56.

43. Chen P, Harcum SW. 2006. Effects of elevated ammonium on glycosylation gene expression in CHO cells. *Metab Eng.* **8**: 123-32.

44. Jaques AJ, Opdenakker G, Rademacher TW, Dwek RA, Zamze SE. 1996. The glycosylation of Bowes melanoma tissue plasminogen activator: lectin mapping, reaction with anti-L2/HNK-1 antibodies and the presence of sulphated/glucuronic acid containing glycans. *Biochem J.* **316**: 427-37.

45. Harazono A, Kawasaki N, Kawanishi T, Hayakawa T. 2004. Site-specific glycosylation analysis of human apolipoprotein B100 using LC/ESI MS/MS. *Glycobiology*. **15**: 447-62.

46. Ackermann M, Marx U, Jager V. 1995. Influence of cell-derived and media-derived factors on the integrity of a human monoclonal-antibody after secretion into serum-free cell-culture supernatants. *Biotechnol. Bioeng.* **45**: 97–106.

47. Gawlitzek M, Conradt HS, Wagner R. 1995. Effect of different cell culture conditions on the polypeptide integrity and N-glycosylation of a recombinant model glycoprotein. *Biotechnol Bioeng.* **46**: 536–544.

48. Hooker AD, Goldman MH, Markham NH, James DC, Ison AP, Bull AT, Strange PG, Salmon I, Baines AJ, Jenkins N. 1995. N-glycans of recombinant interferon-

change during batch culture of Chinese hamster ovary cells. *Biotechnol. Bioeng.* **48**: 639-48.

49. Neusüß C, Demelbauer U, Pelzing M. 2005. Glycoform characterization of intact erythropoietin by capillary electrophoresis-electrospray-time of flight-mass spectrometry. *Electrophoresis*. **26**: 1442-50.

50. Wang Y, Wu SL, Hancock WS. 2006. Monitoring of glycoprotein products in cell culture lysates using lectin affinity chromatography and capillary HPLC coupled to electrospray linear ion trap-Fourier transform mass spectrometry (LTQ/FTMS). *Biotechnol Prog*. **22**: 873-80.

51. Spellman MW, Basa LJ, Leonard CK, Chakel JA, O'Connor JV, Wilson S, van Halbeek H. 1989. Carbohydrate structures of human tissue plasminogen activator expressed in Chinese hamster ovary cells. *J Biol Chem.* **264**: 14100-11.

52. Eaton LC. 1995. Host cell contaminant protein assay development for recombinant biopharmaceuticals. *J. Chromatogr. A*. **705**: 105–114.

53. Rathore AS, Sobacke SE, Kocot TJ, Morgan DR, Dufield RL, Mozier NM. 2003. Analysis for residual host cell proteins and DNA in process streams of a recombinant protein product expressed in *Escherichia coli* cells. *J. Pharm. Biomed. Anal.* **32**: 1199-211.

54. Krawitz DC, Forrest W, Moreno GT, Kittleson J, Champion KM. 2006. Proteomic studies support the use of multi-product immunoassays to monitor host cell protein impurities. *Proteomics.* **6**: 94-110.

55. Hermann T. 2004. Using functional genomics to improve productivity in the manufacture of industrial biochemicals. *Curr. Opin. Biotechnol.* **15**: 444-8.

56. Lee SY, Lee DY, Kim TY. 2005. Systems biotechnology for strain improvement. *Trends Biotechnol.* **23**: 349-58.

57. Park JT, Bradbury L, Kragl FJ, Lukens DC, Valdes JJ. 2006. Rapid optimization of antibotulinum toxin antibody fragment production by an integral approach utilizing RC-SELDI mass spectrometry and statistical design. *Biotechnol Prog*. **22**: 233-40.

58. Aldor IS, Krawitz DC, Forrest W, Chen C, Nishihara JC, Joly JC, Champion KM. 2005. Proteomic profiling of recombinant *Escherichia coli* in high-cell-density fermentations for improved production of an antibody fragment biopharmaceutical. *Appl Environ Microbiol*. **71**: 1717-28.

59. Champion KM, Nishihara JC, Aldor IS, Moreno GT, Andersen D, Stults KL, Vanderlaan M. 2003. Comparison of the *Escherichia coli* proteomes for

recombinant human growth hormone producing and nonproducing fermentations. *Proteomics*. **3**: 1365-73.

60. Champion KM, Nishihara JC, Joly JC, Arnott D. 2001. Similarity of the *Escherichia coli* proteome upon completion of different biopharmaceutical fermentation processes. *Proteomics*. **1**: 1133-48.

61. Han MJ, Jeong KJ, Yoo JS, Lee SY. Engineering *Escherichia coli* for increased productivity of serine-rich proteins based on proteome profiling. *Appl Environ Microbiol*. **69**: 5772-81.

62. Jurgen B, Lin HY, Riemschneider S, Scharf C, Neubauer P, Schmid R, Hecker M, Schweder T. 2000. Monitoring of genes that respond to overproduction of an insoluble recombinant protein in Escherichia coli glucose-limited fed-batch fermentations. *Biotechnol Bioeng.* **70**: 217-24.

63. Kim YH, Park JS, Cho JY, Cho KM, Park YH, Lee J. 2004. Proteomic response analysis of a threonine-overproducing mutant of *Escherichia coli*. *Biochem J*. **381**: 823-9.

64. Raman B, Nandakumar MP, Muthuvijayan V, Marten MR. 2005. Proteome Analysis to Assess Physiological Changes in *Escherichia coli* Grown under Glucose-Limited Fed-Batch Conditions. *Biotechnol Bioeng.* **92**: 384-92.

65. Rinas U. 1996. Synthesis rates of cellular proteins involved in translation and protein folding are strongly altered in response to overproduction of basic fibroblast growth factor by recombinant *Escherichia coli*. *Biotechnol. Prog.* **12**: 196-200.

66. Wang Y, Wu SL, Hancock WS, Trala R, Kessler M, Taylor AH, Patel PS, Aon JC. 2005. Proteomic profiling of *Escherichia coli* proteins under high cell density fed-batch cultivation with overexpression of phosphogluconolactonase. *Biotechnol. Prog*. **21**: 1401-11.

67. Choi JH, Lee SJ, Lee SJ, Lee SY.2003. Enhanced production of insulin-like growth factor I fusion protein in *Escherichia coli* by coexpression of the down-regulated genes identified by transcriptome profiling. *Appl. Environ. Microbiol.* **69**: 4737-42.

68. Gill RT, DeLisa MP, Valdes JJ, Bentley WE. 2001. Genomic analysis of high-cell-density recombinant *Escherichia coli* fermentation and "cell conditioning" for improved recombinant protein yield. *Biotechnol. Bioeng*. **72**: 85-95.

69. Haddadin FT, Harcum SW. 2005. Transcriptome profiles for high-cell-density recombinant and wild-type *Escherichia coli*. *Biotechnol Bioeng*. **90**: 127-53.

70. Harcum SW, Haddadin FT. 2006. Global transcriptome response of recombinant *Escherichia coli* to heat-shock and dual heat-shock recombinant protein induction. *J Ind Microbiol Biotechnol.* **33**: 801-14.

71. Polen T, Rittmann D, Wendisch VF, Sahm H. 2003. DNA microarray analyses of the long-term adaptive response of *Escherichia coli* to acetate and propionate. *Appl Environ Microbiol*. **69**: 1759-74.

72. Wei Y, Lee JM, Richmond C, Blattner FR, Rafalski JA, LaRossa RA. 2001. High-density microarray-mediated gene expression profiling of *Escherichia coli*. *J. Bacteriol.* **183**: 545-56.

73. Yoon SH, Han MJ, Lee SY, Jeong KJ, Yoo JS. 2003. Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. *Biotechnol. Bioeng*. **81**: 753-67.

74. Lee KH. 2001. Proteomics: A Technology-Driven and a Technology-Limited Discovery Science. *Trends Biotechnol.* **19**: 217-222.

75. Aggarwal K, Choe LH, Lee KH. 2006. Shotgun proteomics using the iTRAQ isobaric tags. *Brief. Funct. Genomic Proteomic.* **5**: 112-120.

CHAPTER 4

SILENT MUTATIONS RESULTS IN HLYA HYPERSECRETION BY REDUCING INTRACELLULAR HLYA PROTEIN AGGREGATES

4.1 Preface

This chapter is adapted from: Gupta, P., and Lee, K.H. 2008. Silent mutations result in HlyA hypersecretion by reducing intracellular HlyA protein aggregates. *Biotechnology and Bioengineering* 101: 967-974. It describes the differences between the parent and a hypersecreter *E. coli* strain, containing a synonymous rare codon cluster in the gene of interest. The results suggest that production of high levels of secreted proteins requires a balance between translation and secretion rate.

4.2 Abstract

Escherichia coli is one of the most widely used hosts for the production of recombinant proteins. Extracellular protein secretion has the advantage of reducing protein aggregation and simplifying downstream purification. The introduction of five rare codons in a specific region of α-hemolysin (*hly*A) gene previously was shown to result in eight-fold improvement in secretion of HlyA via the hemolysin (Type-I) pathway. Here we investigate the biological basis for the observed phenomenon that the translation rate of HlyA protein may be related to the ability to secrete higher levels of HlyA via the Type-I pathway. A detailed comparative analysis between a hypersecreter mutant strain (hly-slow) and a control strain (hly-parent) shows a significant decrease (by ~ 50%) in the intracellular level of HlyA protein in the hly-slow strain relative to the hly-parent strain. Nearly 100 % of the intracellular HlyA

protein exists in the inclusion body fraction in both the strains. These results demonstrate the importance of synonymous codon changes in the context of improving HlyA secretion yield via Type-I pathway and further illustrate that production of high levels of secreted proteins appears to require a balance between translation and secretion rate.

4.3 Introduction

The annual global market for biopharmaceuticals is estimated at more than $30 billion and products derived from *Escherichia coli* represent nearly 39% of all the biopharmaceuticals made in the US and EU [1]. Intracellular expression of proteins in the cytoplasm often results in the formation of inclusion bodies, which are not functional and difficult to process, thereby increasing production cost and reducing yield [2]. Secretion of heterologous protein in the periplasm can simplify purification processes but high levels of transport are difficult to sustain and the export mechanisms are not well understood. Inclusion bodies may also form in the periplasm as a consequence of over-expression of recombinant proteins [3].

From a biotechnology perspective, heterologous protein secretion into the extracellular medium is desirable because the concentration of the protein of interest remains low, thus minimizing protein aggregation. Additionally, the extracellular environment can be controlled to provide optimal osmolarity and pH for protein folding and stability. However there have been relatively few successful attempts to secrete proteins into the extracellular medium, some of which have been reported [4-8]. Efforts to alleviate bottlenecks in secretion pathways via metabolic engineering do not often produce the desired phenotype, primarily because of the low efficiency and high specificity of most secretion systems and an incomplete understanding of their mechanisms. While

E. coli exports several types of proteins to the periplasm, secretion pathways to the extracellular medium are often specific for only a few proteins. Understanding and controlling translocation mechanisms enables better control over the secretion systems for the production of recombinant proteins.

The Type-I secretion or hemolysin pathway is one of the simplest secretion systems and has the ability to secrete recombinant proteins directly past the outer membrane. It is responsible for the translocation of a 107 kDa α-hemolysin protein (HlyA) directly from the cytoplasm of pathogenic *E. coli* strains to the extracellular medium [9]. The Type-I secretion machinery is composed of three membrane-associated proteins: HlyB, HlyD, and TolC [10]. The use of the hemolysin pathway to secrete some recombinant proteins has been shown to produce efficiencies of up to 3-5% of total cellular protein accumulating in the medium, which corresponds to approximately 2.3 mg/L/OD [4].

Previous studies have established that the protein synthesis rate is an important consideration for secretion efficiency [6,11]. Of the number of factors that can affect protein synthesis rate, the effect of plasmid copy number, translation initiation region, and differences in codon usage on protein secretion have been studied in greater detail. Recent studies have shown that *E. coli* transformed with plasmids with different copy numbers exhibited altered export of human proinsulin to the periplasm [12]. In this study, it was found that the use of plasmids of moderate copy number (15-60) versus low copy number (11) increased export by two-fold, but the study did not include a high copy number plasmid. Studies on the export of recombinant proteins to the periplasm [11] revealed that a change in the sequence of the translation initiation region (TIR) affects the rate of translocation by the Sec pathway. However, an

increase in protein expression levels (brought about by differences in the strength of TIR) did not uniformly increase the amount of protein secreted. Rather, the optimal TIR (and presumably, expression level) varied for each protein tested. The investigators speculated that the change to the first few codons in the gene sequence that were made might affect the structure of the mRNA resulting in enhanced binding and translation by the smaller ribosomal subunit. Nevertheless, this study demonstrated that an optimum translational level exists to achieve high-level secretion of each heterologous protein, and outside of this optimum level, secretion levels drop off precipitously.

We previously reported that synonymous codon changes (specifically changing abundant codons to rare codons) in a specific region of *hly*A gene resulted in an eight-fold improvement in active HlyA secretion relative to the parent strain [6]. The improved HlyA secretion rate was quantified by liquid blood lysis assay [6]. Rare codons are defined as those codons whose corresponding tRNA concentration is less than 1% of the total tRNA concentration [as tabulated in [13]]. The resulting strain, hly-slow, contained five synonymous codon changes in the *hly*A gene and was engineered to have a predicted 37% decrease in HlyA translation rate [14]. The study suggested that an optimal translation rate is required for maximum secretion of HlyA protein via Type-I secretion pathway. Here, we perform a comparative analysis of the hly-parent and hly-slow strains to better understand the enhanced secretion phenotype in the hly-slow strain. Specifically, we show that the presence of rare codons in the hly-slow strain results in less HlyA aggregates relative to the hly-parent strain and consequently more active secreted HlyA in the hly-slow strain.

4.4 Materials and Methods

4.4.1 Plasmids and strains

All plasmids used in this study are shown in Table 4.1. Plasmids pWAM1097 and pWAM716 [15,16] were obtained from Rodney Welch (University of Wisconsin-Madison). The strains used in this study are W3110-based and are listed in Table IB. 6X-His tagged *hly*A gene was constructed as follows. First the *hly*A gene from pWAM1097 was PCR amplified with specific forward primer 5'-CAGTGCTAGCACATCATC ACCATCACCATATGCCAACAATAACC GC- 3' containing 6X-His tag and *Nhe*I restriction enzyme site and reverse primer 5'-GCAGAAGCTTTTATGCTGATGCTGTCAAAG - 3' incorporating *Hind*III restriction enzyme site. Second, this PCR product was double digested with *Nhe*I and *Hind*III restriction enzymes and ligated into the same sites of pWAM1097 vector to form pWAM1097_HisHlyA plasmid. phlyslow plasmid was created by site directed mutagenesis of pWAM1097_HisHlyA.

4.4.2 Liquid blood lysis assay

The liquid blood lysis assay was adapted from several previously reported hemolysis assays [17-19]. HlyA secreting cells were grown to mid-logarithmic phase in tryptone water (1% tryptone, 0.5% NaCl) at 37°C with shaking at 250 rpm. Cultures were diluted to OD_{600} = 0.1 and grown for 1–3 hr at 37°C with shaking at 250 rpm. Cells were again diluted to OD_{600} = 0.1 and centrifuged. The supernatant was removed and serially diluted over three orders of magnitude. Sheep erythrocytes (Hardy Diagnostics) were washed at least three times in 0.9% sodium chloride by centrifugation to remove hemoglobin from lysed cells. A 4% sheep erythrocyte suspension was made in 0.9% sodium chloride. Hemolysis was monitored in 96-well plates. A reaction buffer was made consisting of 0.9% sodium chloride with 10 mM

Table 4.1. Table of plasmids and strains used in this study; (A) shows plasmids used in the study and their sources (B) shows strains used and their origin.

(A) Plasmid	**Genes**	**Copy no.**	**Resistance**	**Source**
pWAM1097	hlyCA	High	Ampicillin	(Felmlee et al., 1985a)
pWAM716	hlyBD	Low	chloramphenicol	(Felmlee et al., 1985b)
pWAM1097_HisHlyA	hlyC, 6X-His tagged hlyA	High	Amplicillin	This study
phlyslow	hlyC, 6X-His tagged hlyA	High	Ampicillin	This study (site-directed mutagenesis of pWAM1097_HisHlyA)

(B) Strain	**Plasmid**	**Origin**
W3110	None	
hly-parent	pWAM1097_HisHlyA, pWAM716	W3110-based
hly-slow	phlyslow, pWAM716	W3110-based

calcium chloride. Each well contained 80 µL of the diluted supernatant, 100 µL reaction buffer, and 20 µL of 4% sheep erythrocytes. Diluted supernatants were assayed in triplicate. Plates were mixed using the Versamax microplate reader (Molecular Devices) for 15 sec to begin the reaction, and then incubated at 37°C. At one hour intervals, the plates were mixed for 15 sec, followed by an OD_{530} measurement. Undiluted supernatant and tryptone water were used as controls. Hemolysis calculations were based on the differences between diluted supernatant samples and controls. The predicted dilution required for 50% hemolysis was calculated by fitting a line to the slope of the lysis curve.

4.4.3 Vancomycin assay

This protocol has been adopted [20]. The cells expressing the Type-I secretion system were grown in 125 mL culture flasks at 37°C in Luria-Bertani (LB) media supplemented with ampicillin (75 µg/mL) and chloramphenicol (85 µg/mL) until they reached OD_{600} = 1. Thereafter, aliquots of cells were taken and incubated at 37°C for 30 min with mild shaking (200-250 rpm) in the presence of different concentrations of vancomycin antibiotic. After vancomycin treatment, cells were plated on LB-agar plates in the absence of vancomycin and supplemented with ampicillin (75 µg/mL) and chloramphenicol (85 µg/mL). Survival was reported as a percentage of the number of colonies formed by control samples previously incubated in the absence of vancomycin.

4.4.4 Site-directed mutagenesis

Primers were obtained from IDT Technologies. The Stratagene QuikChange site-directed mutagenesis kit was used with a BIORAD *icycler*™ thermocycler.

Confirmation of the desired changes was obtained by sequencing of purified plasmids at DNA services, Cornell University, Ithaca, NY.

4.4.5 Quantitative real time reverse transcription polymerase chain reaction (qRT-PCR)

*hly*A mRNA level was evaluated by comparative real-time RT-PCR utilizing TaqMan® One-Step RT-PCR Master Mix Reagents Kit (Applied Biosystems) and ABI 7900HT Sequence Detection System (Applied Biosystems). HlyA secreting cells were grown in 125 ml culture flasks at 37°C in Luria- Bertani (LB) media supplemented with ampicillin (75 μg/mL) and chloramphenicol (85 μg/mL) with shaking at 250 rpm and total RNA was extracted from the cells using Qiagen RNeasy kit (Qiagen), as per the manufacturer's protocol. 10 ng of total RNA was used for each qRT-PCR reaction and 16S rRNA gene was used as an internal control. The *hly*A specific forward primer 5'-GGTATTCGGCACAGCA GAGAA -3' and the reverse primer 5'-GTCTAATTGTGGTGC AAAGATAGTCACT -3' and 16S rRNA specific forward primer 5'-CCAGCAGCCGCGGTAA T -3'and the reverse primer 5'-TGCGCTTTACGCCCAGTAAT -3' were used in these studies. The TaqMan® Probe with 6-carboxyfluorescein (6-FAM) as the reporter dye and tetramethyl-6-carboxyrhodamine (TAMRA) as the quencher was used and the 16S rRNA probe sequence was 5'-CCGATTAACGCTTGCACCCTCCG -3' and *hly*A probe sequence was 5'-CTCATTGGCC TCACCGAACGGG -3'. The threshold cycle for each amplification curve was calculated using SDS software, version 2.1 (Applied Biosystems).

4.4.6 Protein fractionation

Single colonies of the cells transformed with the appropriate plasmids were grown in 125 ml culture flasks at 37°C in Luria- Bertani (LB) media supplemented with

ampicillin (75 μg/mL) and/or chloramphenicol (85 μg/mL). Aliquots of cells were harvested at similar OD_{600} for all the samples and pelleted by centrifugation for 10 min at 4°C and 4000 × g. 500 μL of the supernatant fraction was transferred to another eppendorf tube and mixed with two volumes of ice-cold ethanol. The tubes were left overnight at -20°C for protein precipitation and centrifuged next day at 5000 rpm for 10 min at 4°C. The resulting protein pellets were resuspended in 20 μL of 1X phosphate buffered saline (PBS) and analyzed using 1-D SDS-PAGE and western blotting. For intracellular protein fractionation from whole cell pellets, the pellets were dissolved in 100 μL of SDS sample buffer, heated at 94°C for 10 min, and analyzed using SDS-PAGE and western blotting.

4.4.7 Inclusion body preparation

The protocol has been adopted [21]. Five OD_{600} units of bacterial cells were resuspended in 400 μL of 50 mM Tris (pH 8.0)/1 mM EDTA (hypotonic buffer), brought to 0.5 mg/mL lysozyme, and incubated for 5 min at room temperature (RT). $MgCl_2$ was added to a final concentration of 5 mM, and the mixture was treated with 30 units of DNase I for 5 min. Sonication was then carried out at 4°C. Triton X-100 was added to 0.5%, and the sample was centrifuged at 4°C for 15 min at 14,000 × g. The pellet fraction was washed once with 500 μL of buffer A (5% acetonitrile/0.1% formic acid) and solubilized with 450 μL of 10M urea/100mM Tris (pH 7.4)/1mM Tris (2-carboxyethyl) phosphine at RT for 30 min. Both soluble and inclusion body fractions were then analyzed using SDS-PAGE and western blotting.

4.4.7 Western Analysis

Supernatant fractions and intracellular protein fractions (soluble and inclusion body) were resolved by SDS-PAGE (12% w/v) using Tris–HCl and immunoblotted. Mouse

anti-6X-His (1:3,000; Sigma) and alkaline phosphatase conjugated goat anti-mouse IgG antibody (1:30,000; Sigma) were used as the primary and secondary antibodies respectively, for the detection of HlyA protein. Bound antibodies were detected using enhanced chemifluorescence (ECF) substrate (GE Amersham Biosciences) following the manufacturer's instructions and imaged using a FLA-3000 Fujifilm scanner. Quantitative analysis of the western blots was done using ImageMaster 2D Platinum Software v5.0 (GE Amersham Biosciences).

4.5 Results

There are five nucleotide differences in the *hly*A gene in the hly-slow strain relative to the hly-parent strain [6]. The five nucleotide changes incorporate rare codons in the hly-slow sequence, while keeping the same amino acid sequence. To study the basis for increased secretion of HlyA protein in the hly-slow strain (hly-S) compared to hly-parent strain (hly-P), we considered four possible contributing factors: higher expression of *hly*A mRNA, higher expression of secretion machinery, more favorable external environment, and higher synthesis of protein per mRNA.

Quantitation of *hly*A mRNA in the hly-P and hly-S was performed using comparative real time RT-PCR. The 16S rRNA gene sequence was used as an endogenous control for this experiment. As seen from the amplification plot (Fig. 4.1), the threshold cycle of the 16S rRNA and *hly*A mRNA for both hly-P and hly-S overlap with each other in the exponential phase of amplification, suggesting that there is no difference in the *hly*A mRNA expression in the two strains. To test for higher expression of transport machinery in the hly-S, we relied on the observation that cells that simultaneously express hlyBD and hlyCA are hypersensitive to vancomycin, an antibiotic that is relatively inactive against gram-negative bacteria [20]. An increase in sensitivity to transport machinery in a given strain. Figure 4.2 shows percentage viable cells for

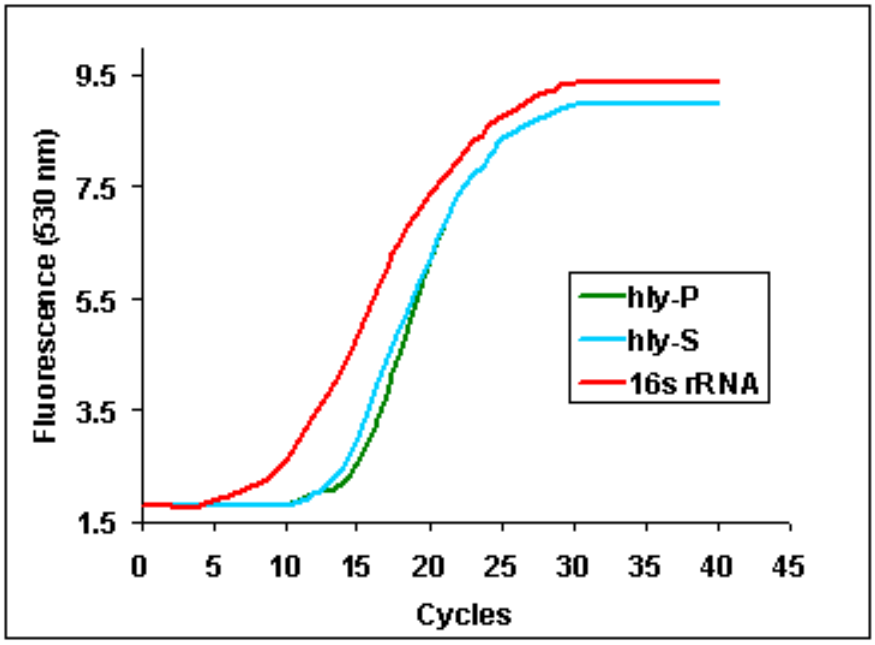


Figure 4.1. Real time RT-PCR to quantify the amount of *hly*A mRNA in hly-parent (hly-P) and hly-slow (hly-S) strains. 16S rRNA gene was used as an endogenous control. The threshold cycle for each amplification curve was calculated using SDS software, version 2.1 (Applied Biosystems) and the fold change difference was calculated relative to the endogenous control. The fold change in *hly*A mRNA expression in the two strains was found out to be 0.95.

vancomycin would be consistent with elevated expression levels of functional Type-I both hly-P and hly-S at different vancomycin concentrations. The data is consistent with no significant difference in the number of transporters in the hly-S relative to hly-P.

An eight fold increase in secreted HlyA protein activity in hly-S relative to hly-P may also result from the over-expression of specific chaperones and/or proteases or any other proteins present in the extracellular media, which might affect HlyA protein activity or stability. To test whether there is a significant difference between the extracellular proteome of the hly-S as compared to the hly-P, we profiled the supernatant fractions for both the strains (Fig. 4.3) to identify changes in the secreted proteins. The results are qualitatively similar, suggesting that the two strains have similar extracellular proteome profile. In this experiment we observed a slight decrease in HlyA (identified by mass spectrometry) expression in the supernatant fraction of hly-P relative to hly-S, as measured by total protein stain (Fig. 4.3). The inconsistent correlation between HlyA protein expression and HlyA activity in hly-S relative to hly-P may be attributed to different levels of acylated (active) and non-acylated (inactive) HlyA in the supernatant fraction of the two strains. Conversion of pro-HlyA to the acylated-HlyA (hemolytically active HlyA) takes place in the cytoplasm of *E. coli* and is mediated by HlyC, however HlyC is not required for the secretion of pro-HlyA [22]. Hence, the slower translation rate of HlyA in the hly-slow strain might result in increased secretion of acylated HlyA relative to the total HlyA secretion, resulting in increased activity. Also, the lack of availability of an HlyA antibody makes the precise quantitation of HlyA expression difficult. Intracellular HlyA protein quantitation was performed to investigate whether there may be more HlyA protein synthesis per mRNA molecule in the hly-slow strain relative to the hly-

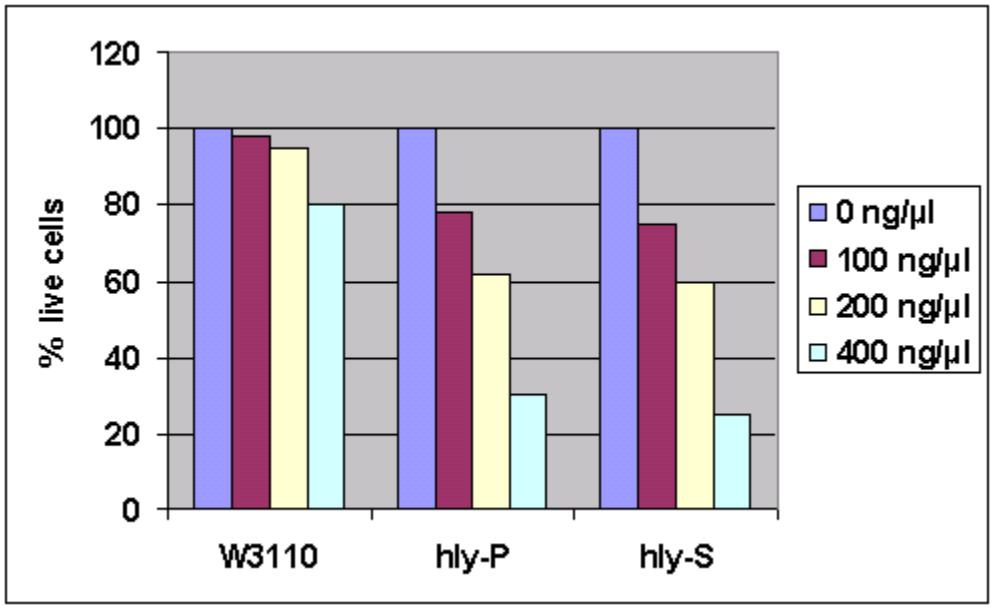


Figure 4.2. Cell viability in the presence of different concentrations of vancomycin antibiotic test whether enhanced secretion phenotype is a result of increased expression of secretion machinery. The colored bars indicate percentage viable cells of hly-parent (hly-P) and hly-slow (hly-S) strains at different concentration of vancomycin relative to the cells grown in the absence of vancomycin. W3110 is used as a control strain.

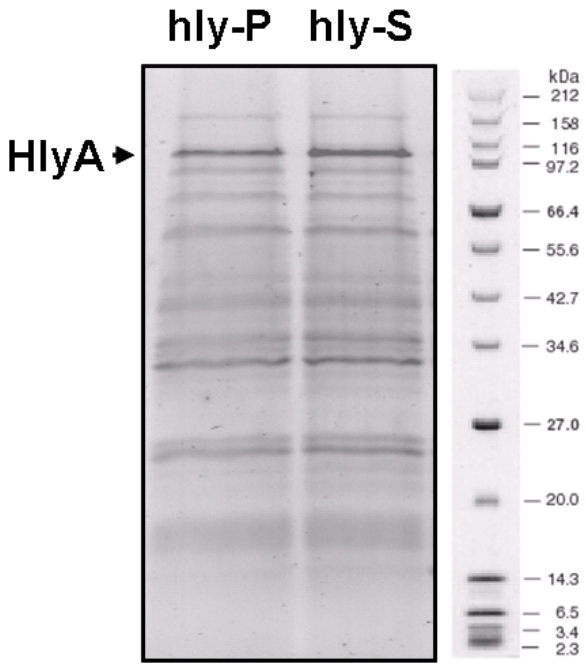


Figure 4.3. Extracellular proteome profile of hly-parent (hly-P), and hly-slow (hly-S) using 1-D SDS-PAGE. The supernatant fractions were collected at similar cell optical density (OD_{600}). HlyA protein band (~110 KDa) was confirmed by mass spectrometry analysis.

parent strain. This experiment was done in hly-parent and hly-slow strains with no transport machinery (hlyBD$^-$) to decouple the effects of intracellular protein production from secretion. Quantitation of HlyA protein in the cell pellet fractions by ECF western analysis (Fig. 4.4) indicates a decrease of about 50% in the intracellular level of HlyA protein in the secretion deficient slow strain, hly-S (BD$^-$), relative to the secretion deficient parent strain, hly-P (BD$^-$), and is consistent with the 37% predicted decrease in translation rate of *hly*A gene in hly-slow strain. Soluble and inclusion body fractions of HlyA were also measured (Fig.4.4). Nearly all of the HlyA protein is present in the inclusion body fraction of the secretion deficient cells, consistent with HlyA protein being prone to intracellular protein aggregation. Further, HlyA quantitation in the inclusion body fraction by ECF western analysis indicates a decrease of about 50% in hly-S (BD$^-$) relative to hly-P (BD$^-$). We also performed intracellular HlyA quantitation in secretion competent cells and observed a similar decrease in hly-slow (BD$^+$) strain relative to hly-parent (BD$^+$) strain (Fig. 4.5).

A decrease in HlyA protein synthesis rate may result from: (a) enhanced protein degradation because of ribosome stalling at the rare codon cluster (SsrA response), (b) a significant change in mRNA secondary structure of *hly-slow* mRNA and/or, (c) slower translation due to the presence of a rare codon cluster. It has been suggested in the literature that the presence of a stalled ribosome on the mRNA, either due to an unwanted stop signal or due to the presence of rare codons, can elicit an SsrA protein degradation response in the cell [23]. To test for the effect of the SsrA response, an SsrA-defective (W3110 Δ*smpB-1, ssrA::kan*) strain was transformed with appropriate plasmids to create hly-parent (*ssr*A$^-$) and hly-slow (*ssr*A$^-$) strains respectively. Both intracellular HlyA protein quantitation (in the absence of transport machinery, as indicated by hlyBD$^-$) (Fig. 4.5) and HlyA secretion assay (Fig. 4.5) was performed.

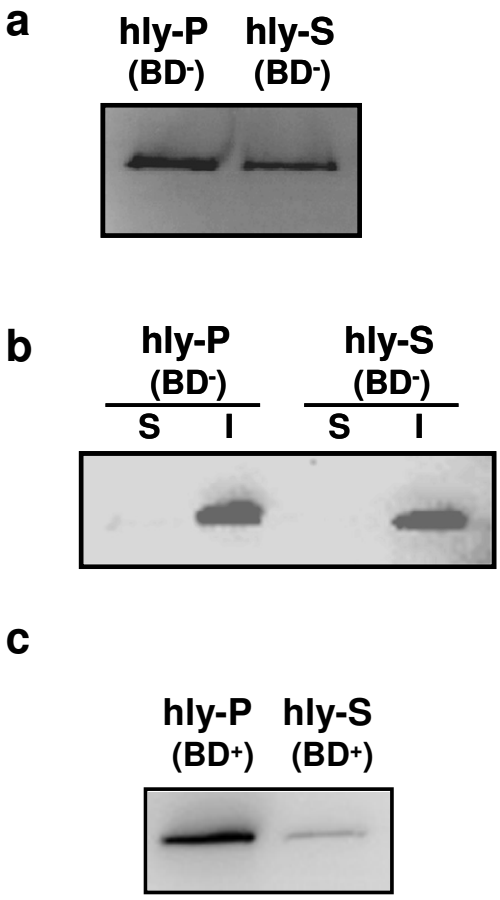


Figure 4.4. Normalized western blot of intracellular HlyA protein in (a) whole cell pellets of secretion deficient strains (BD^-), (b) Soluble (S) and inclusion body (I) fractions of secretion deficient parent strains, and (c) whole cell pellets of secretion competent cells (BD^+). An equivalent number of cells were harvested for each experiment.

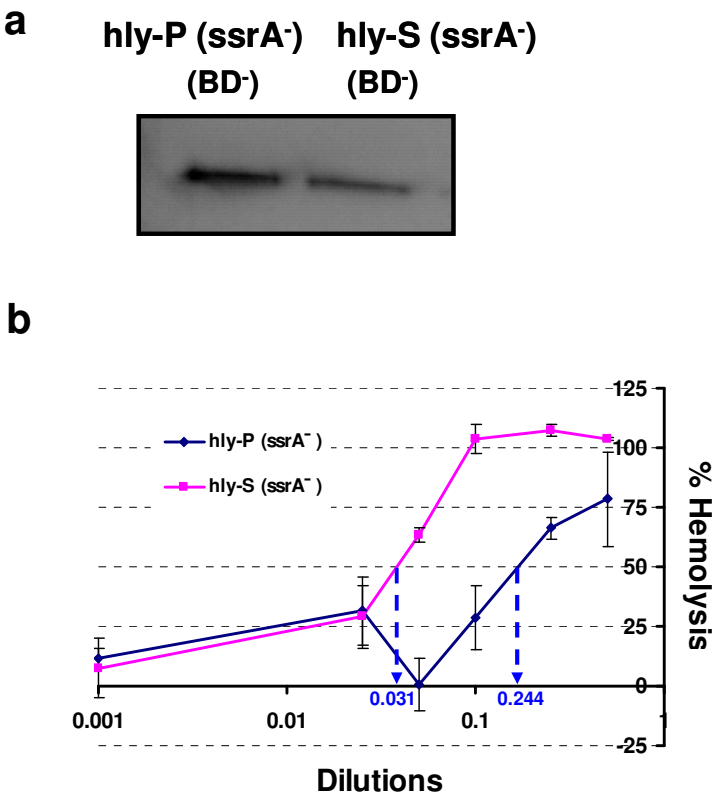


Figure 4.5. Intracellular HlyA quantitation and secretion studies in *ssr*A (-) background. (a) Normalized western blot of HlyA protein in whole cell pellets in the secretion deficient parent strain, hly-P (BD^-) and secretion deficient slow strain, hly-S (BD^-). (b) Liquid blood assay for secretion competent hly-P and hly-S strains. Each reading was performed in triplicate and error bars show the standard deviation of triplicate measurements. Fold change calculations were made as follows. The supernatant dilution corresponding to 50% hemolysis was calculated (numbers in bold marked by red dashed arrows) from the lysis curve and secretion fold change was calculated by dividing this dilution.

Quantitative western analysis suggests an approximately 50% decrease in the intracellular levels of HlyA protein in hly-slow strain relative to hly-parent and liquid blood assay indicates approximately a 7.87 fold increase in active HlyA secretion in hly-slow strain relative to parent. These results are consistent with the previous observations in hly-parent (*ssr*A^+) and hly-slow (*ssr*A^+) strains, suggesting that the observed phenotypes (decrease in intracellular HlyA production and increase in secreted HlyA activity) are not affected by the SsrA response in the cell.

mRNA secondary structure analysis of *hlyA-slow* mRNA and *hly-parent* mRNA was also performed using the mRNA secondary structure prediction program, MFOLD [24]. Fig. 6 shows the difference in the predicted stem loop structures of the two mRNAs. A small Gibb's free energy difference of 1.09 kcal/mol between the two structures. suggests that the translation of the HlyA protein in the hly-slow strain is not affected by the change in the *hly*A mRNA secondary structure. Hence the decreased intracellular HlyA expression in hly-slow strain is likely a result of slower HlyA translation rate.

4.6 Discussion

Differences in codon usage between the source organism and the production host have been observed to affect the translation rate of recombinant proteins because codon usage directly correlates with the abundance of codon-specific tRNAs available within the cell [26]. Generally, a synonymous substitution of the codons in the gene of interest to more commonly used codons in the host cell alleviates this problem [27]. However, it has been hypothesized that a significant increase in protein expression can overwhelm any given secretory pathway [11]. Thus, to achieve high levels of

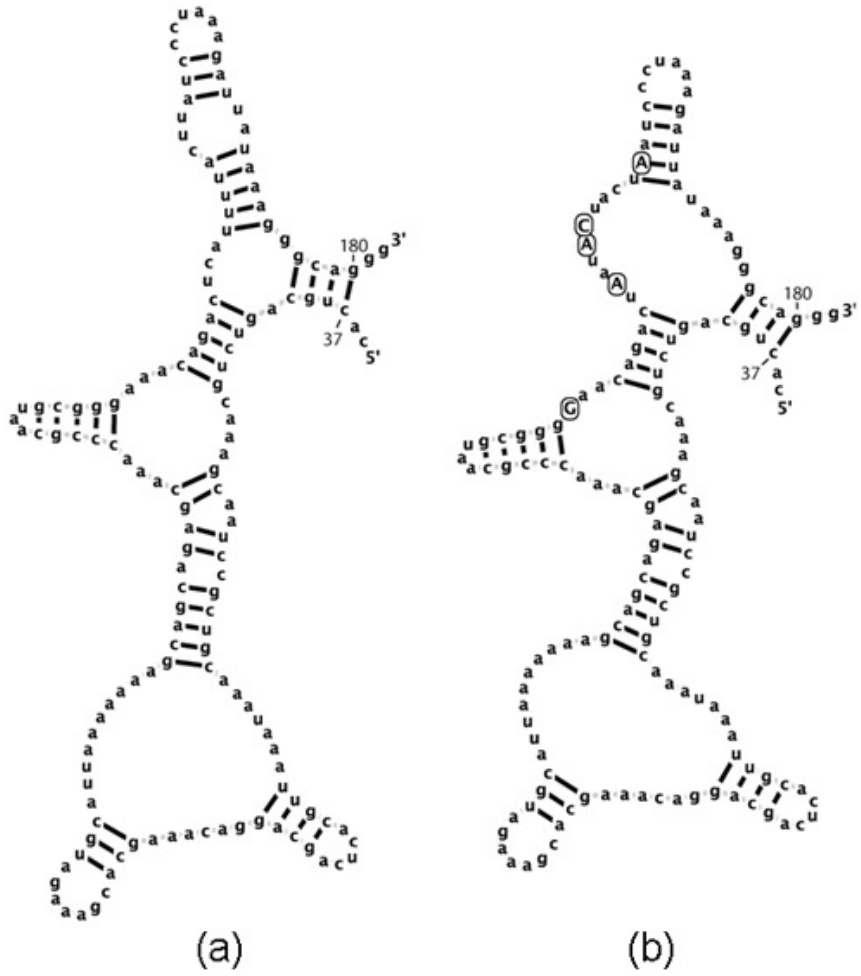


Figure 4.6. The local stem loop mRNA structure of *hly*A gene in (a) hly-parent, and (b) hly-slow strain, as predicted by the program MFOLD [23]. The figures were generated using the program jViz.Rna v1.77 [25]. The differences between the two gene sequences are circled and the predicted Gibbs free energy difference between the two stem-loop structures is approximately +1.09 kcal/mol.

expression and complete processing of the precursor, the translation rate of a heterologous protein destined for the secretion pathway requires optimization, not simply maximization. We previously demonstrated that a rare codon cluster in the target protein (HlyA) gene sequence can lead to an enhanced secretion capability via Type-1 secretion machinery [6]. From a mathematical model of translation [14], the new *hly*A gene sequence has a predicted translation rate decrease of 37% and empirically this was achieved by changing five of the adjacent codons to rare codons in a specific region of *hly*A gene but without altering the amino acid sequence. That observation suggests that a balanced synthesis rate is important because too many rare codons may not permit enough expression, while too few rare codons may reduce the secretion rate by overwhelming the secretory apparatus in the cell. We investigated different factors that may account for the enhanced secretion phenotype in the hly-slow strain and our data suggests that the observed phenotype is the result of decreased total HlyA protein production relative to the hly-parent strain. We also showed that most of the intracellular HlyA exists in the form of inclusion body fraction in secretion incompetent cells. This observation is consistent with the idea that intracellular HlyA is mostly unfolded because of the lack of Ca^{+2} ions, which are thought to bind to the RTX repeats in HlyA protein, rich in glycine and aspartic acid residues, and initiate HlyA folding [28]. We hypothesize that the HlyA protein aggregates might interact with the Type-I translocase to form unwanted protein complexes, thereby hindering the transport process.

Another line of hypothesis suggests that there might be secretion competent and secretion incompetent pools of substrate protein inside the cell and the introduction of rare codons in a specific stretch of the gene of interest can affect the balance of these pools and the secretion flux. That is, the type and position of the rare codons

introduced may result in populations of different folded/unfolded conformations, favoring secretion. Recent studies have shown that the introduction of synonymous codon changes in specific regions of the gene of interest can affect the substrate specificity of a polypeptide [29,30]. It has been shown that synonymous codon substitutions in chloramphenicol acetyltransferase (CAT) gene can affect the CAT protein activity during *in vitro* translation [29]. Also in a recent study, it was shown that a certain synonymous codon mutation in the multidrug resistance 1 (*MDR*1) gene in HeLa cells changed the substrate specificity of the *MDR*1 gene product P-glycoprotein [30]. Interestingly, the authors in this study hypothesized that the timing of folding can be affected by interchanging frequently occurring or abundant codons with the rare codons in a specific codon cluster.

To our knowledge, this is the first study which empirically tests the effect of silent mutations on the HlyA secretion phenotype. Overall, this study underlines the importance of synonymous codon changes in the context of improving HlyA secretion yield and may be generalized for the secretion of different recombinant proteins via Type-I pathway. We also hypothesize that the effect of rare codons on the HlyA secretion yield might be a manifestation of not only the HlyA translation rate, but also HlyA folding.

4.7 Acknowledgements

The authors would like to thank Rodney Welch for kindly donating hemolysin plasmids, Matthew DeLisa for donating ssrA deletion mutant, and Robert S. Kuczenski for help with jViz.Rna program. KHL is supported by NYSTAR, NSF, and University of Delaware.

REFERENCES

1. Walsh, G. 2006. Biopharmaceutical benchmarks - 2006. *Nat. Biotechnol.* **24**:769-776.

2. Villaverde A, Carrio MM. 2003. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol. Lett.* **25**:1385-1395.

3. Georgiou G, Telford JN, Shuler ML, Wilson DB. 1986. Localization of inclusion bodies in *Escherichia coli* overproducing beta-lactamase or alkaline phosphatase. *Appl. Environ. Microbiol.* **52**:1157-1161.

4. Blight MA, Holland IB. 1994. Heterologous protein secretion and the versatile *Escherichia coli* haemolysin translocator. *Trend. Biotechnol.* **12**:450-455.

5. Li YY, Chen CX, von Specht BU, Hahn HP. 2002. Cloning and hemolysin mediated secretory expression of a codon-optimized synthetic human interleukin-6 gene in *Escherichia coli*. *Prot. Expr. Purif.* **25**:437- 447.

6. Lee PS, Lee KH. 2005. Engineering HlyA Hypersecretion in *Escherichia coli* Based on Proteomic and Microarray Analyses. *Biotechnol. Bioeng.* **89**:195-205.

7. Sugamata Y, Shiba T. 2005. Improved secretory production of recombinant proteins by random mutagenesis of hlyB, an alpha-hemolysin transporter from *Escherichia coli*. *Appl. Environ. Microbiol.* **71**:656-662.

8. Zhang G, Brokx S, Weiner JH. 2006. Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in *Escherichia coli*. *Nat. Biotechnol.* **24**:100-104.

9. Felmlee T, Pellett S, Lee EY, Welch RA. 1985a. *Escherichia coli* hemolysin is released extracellularly without cleavage of a signal peptide. *J. Bacteriol.* **163**:88-93.

10. Thanabalu T, Koronakis E, Hughes C, KoronakisV. 1998. Substrate-induced assembly of a contiguous channel for protein export from *E. coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *EMBO J.* **17**:6487–6496.

11. Simmons LC, Yansura DG. 1996. Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. *Nat. Biotechnol.* **14**:629-634.

12. Mergulhao FJ, Monteiro GA, Larsson G, Sanden AM, Farewell A, Nystrom T, Cabral JM, Taipa MA. 2003. Medium and copy number effects on the

secretion of human proinsulin in *Escherichia coli* using the universal stress promoters uspA and uspB. *Appl. Microbiol. Biotechnol.* **61**:495-501.

13. Solomovici J, Lesnik T, Reiss C. 1997. Does *Escherichia coli* optimize the economics of the translation process? *J. Theor. Biol.* **185**:511-521.

14. Shaw L, Zia R, Lee KH. 2003. Totally asymmetric exclusion process with extended objects: A model for protein synthesis. *Phys. Rev. E* **68**:021910 (1-17).

15. Thomas WD Jr, Wagner SP, Welch RA. 1992. A heterologous membrane protein domain fused to the C-terminal ATP-binding domain of HlyB can export *Escherichia coli* hemolysin. *J. Bacteriol.* **174**:6771-6779.

16. Felmlee T, Welch RA. 1988. Alterations of amino acid repeats in the *Escherichia coli* hemolysin affect cytolytic activity and secretion. *Proc. Natl. Acad. Sci. USA* **85**:5269-5273.

17. Cortajarena A, Goni FM, Ostolaza H. 2002. His-859 is an essential residue for the activity and pH dependence of *Escherichia coli* RTX toxin alpha-hemolysin. *J. Biol. Chem.* **277**:23223-23229.

18. Jurgens D, Ozel M, Takaisi-Kikuni NB. 2002. Production and characterization of *Escherichia coli* enterohemolysin and its effects on the structure of erythrocyte membranes. *Cell Biol. Intl.* **26**:175-186.

19. Vakharia H, German GJ, Misra R. 2001. Isolation and characterization of *Escherichia coli* tolC mutants defective in secreting enzymatically active alpha-hemolysin. *J. Bacteriol.* **183**:6908–6916.

20. Pimenta AL, Young J, Holland IB, Blight MA. 1999. Antibody analysis of the localisation, expression and stability of HlyD, the MFP component of the *E. coli* haemolysin translocator. *Mol. Gen. Genet.* **261**:122-132.

21. Chapman E, Farr GW, Usaite R, Furtak K, Fenton WA, Chaudhuri TK, Hondorp ER, Matthews RG, Wolf SG, Yates JR, Pypaert M, Horwich AL. 2006. Global aggregation of newly translated proteins in an *Escherichia coli* strain deficient of the chaperonin GroEL. *Proc. Natl. Acad. Sci. USA* **103**:15800-15805.

22. Stanley P, Koronakis V, Hughes C. 1998. Acylation of *Escherichia coli* hemolysin: a unique protein lipidation mechanism underlying toxin function. Microbiol. *Mol. Biol. Rev.* **62**:309-33.

23. Karzai AW, Roche ED, Sauer RT. 2000. The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat Struct Biol*. **7**: 449-455.

24. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**: 3406-3415.

25. Wiese KC, Glen E, Vasudevan A. 2005. jViz.Rna - A Java Tool for RNA Secondary Structure Visualization. *IEEE Transactions on NanoBioscience* **4**:212-218.

26. Ikemura T. 1981a. Correlation between the abundance of *E. coli* transfer RNAs and the occurrence of the respective codon in the protein genes. *J. Mol. Biol.* **146**:1-21.

27. Williams DP, Regier D, Akiyoshi D, Genbauffe F, Murphy JR. 1988. Design, synthesis and expression of a human interleukin-2 gene incorporating the codon usage bias found in highly expressed *Escherichia coli* genes. *Nucleic Acids Res.* **16**:10453-10467.

28. Lilie H, Haehnel W, Rudolph R, Baumann U. 2000. Folding of a synthetic parallel beta-roll protein. *FEBS Lett.* **470**:173-177.

29. Komar AA, Lesnik T, Reiss C. 1999. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett.* **462**:387-391.

30. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. 2007. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**:525-528.

CHAPTER-5

SYNONYMOUS RARE CODON CLUSTER CAN ENHANCE PROTEIN SECRETION BY AFFECTING INTERACTION WITH MOLECULAR CHAPERONES

5.1 Preface

Silent mutations do not change the amino acid composition of the protein product and they have largely been assumed to exert no discernible effect on gene function or phenotype. However, recent studies have hinted that synonymous codon replacement can change protein folding and activity, indicating that protein structure depends on DNA sequence. The following chapter discusses the effect of synonymous rare codon cluster on protein folding and secretion in *E. coli*.

5.2 Abstract

Synonymous codon substitutions can alter protein folding, structure and function; although the mechanism is not well understood. Here, a series of experiments have been described to study the effect of synonymous rare codon cluster on protein folding and secretion via multiple pathways in *E. coli*. Significant improvement in the secretion of various recombinant proteins by the Type-I, Sec and Tat pathways was observed. The analyses revealed that synonymous rare codon cluster at specific sites of beta-lactamase gene (*bla*) can alter folding and subsequent secretion via the Tat pathway. We observed that Bla protein expressed with Tat and Type-I secretion signal peptides interact with elongation factor (EF-Tu), which may prevent it from folding in a secretion competent conformation. Synonymous rare codon substitutions in the hemophore gene (*hasA*) promoted secretion of HasA in a SecB chaperone-independent manner, suggesting slower HasA folding. We conclude that synonymous rare codon

cluster can enhance protein secretion in *E. coli* by affecting interactions with molecular chaperones.

5.3 Introduction

Species-specific disparities in codon usage are frequently cited as the cause for failures in recombinant gene expression by heterologous hosts. Such failures include a lack of expression, expression of a non-functional or insoluble protein, or protein truncation due to proteolysis or premature termination translation [1-3]. These problems may be overcome by either codon optimization of the target genes [4-8] or redesign of genes by substituting the native codons with synonymous codons having similar usage frequencies in the expression host [9,10]. It has also been shown that synonymous codon changes (specifically replacement of abundant codons by rare codons) can significantly improve extracellular secretion in *E. coli* by reducing intracellular aggregates of the recombinant protein [11,12]. These studies suggest that synonymous variations in the nucleotide sequence can alter the elongation rate of the protein of interest and may impact the fate of the polypeptide during or after the process of translation.

In *E. coli,* as well as in eukaryotes, nascent proteins fold co-translationally within the ribosomal tunnel, which is a dynamic environment that influences nascent protein structure [13-16]. Hence, subtle variations in the elongation rate may play a key role in developing secondary structure in the nascent protein and modulate the folding of the translated polypeptide [17]. A recent study has shown that synonymous substitutions that introduce rare codons into regions predicted to contain high frequency codons can affect substrate specificities [18]. Conversely, synonymous codon substitutions that replace rare codons with high frequency codons in regions of slow mRNA translation

can deleteriously affect enzyme activity [19,20]. Thus, contrary to conventional thinking, synonymous codon substitutions may not always be silent; changing codon usage frequency affects protein structure and function, and the frequency with which codons are used imparts vital information for the development of secondary and tertiary protein structure.

Here, we investigated the effect of silent mutations on protein secretion by the Type-I, Sec, Tat and HasABC pathways in *E. coli*. Significant improvement in extracellular and periplasmic secretion of several recombinant proteins was observed by synonymous rare codon engineering. Synonymous rare codon cluster at specific sites of *bla* gene altered Bla folding, activity and subsequent secretion via the Tat pathway. We observed that Bla protein expressed with the Type-I and Tat signal peptides interacts with EF-Tu chaperone and the presence of synonymous rare codon cluster can modulate this interaction to increase or inhibit secretion. These results demonstrate the utility of synonymous codon engineering in improving protein secretion in *E. coli* and reveal that silent mutations modulate the interactions with molecular chaperones and consequently affect protein folding.

5.4 Materials and Methods

5.4.1 Plasmids and strains

All primers, plasmids and strains used in this study are shown in Table 5.1. Plasmids encoding native IL-6 (IL6), codon optimized IL-6 (coIL6), beta-lactamase (Bla), and NYESO1 (ESO1) were fused with the Type-I secretion signal sequence in the following manner. The *il6* gene was PCR amplified with primers #1 and # 2 using pCMV-SPORT6_IL6 plasmid (ATCC number MGC-9215) as the template and the HlyA secretion signal sequence (180 bp) was PCR amplified with primers #2 and #3

using pWAM1097_HisHlyA as the template. These fragments were ligated and cloned Into the *Hind*III and *EcoR*I sites of the pEGFP vector (Clontech, CA, USA). The codon-optimized *il*6 gene was custom synthesized with *Hind* III (5'end) and *Bgl*II (3'end) restriction sites (BioBasic, Canada) and cloned into similar sites in pEGFP. Fusion genes (*il6* and *coil6* fused with the Type-I signal sequence) were PCR amplified with primers # 5 and # 6 (*il6*) and primers #7 and #6 (*coil6*) and cloned into the *Nhe*I and *Hind*III restriction sites of pWAM1097_HisHlyA to create the pWAM1097_IL6 and pWAM1097_coIL6 plasmids. *bla* and *nyeso1* (gift from Ludwig Institute for Cancer Research, NY, USA) genes were PCR amplified with primers #8 and #9 (*bla*), and primers #10 and #11 (*nyeso1*) and cloned into the *Nhe*I and *Bgl*II restriction sites of pWAM1097_IL6 to create pWAM1097_coIL6, pWAM1097_bla and pWAM1097_ESO1 plasmids. Further, the kanamycin (*kan*) resistance marker gene was PCR amplified with primers #12 and #13 and cloned into the *Aat*II and *Sca*I restriction sites in pWAM1097_Bla to create the pWAM1097kan_Bla plasmid. The phasADE plasmid (containing compelete HasA secretion system) was a gift by Cécile Wandersman (Institut Pasteur, France) and pTrc99a-cm_Tat-Bla (containing *bla* gene tagged with ssTorA Tat signal peptide) was a gift from Matthew DeLisa (Cornell University, USA). pTrc99a-cm_Sec-Bla was created by PCR amplifying the precursor *bla* gene (containing the Sec signal sequence) using primers #14 and #15 and then cloning into the *Sac*I and *Hind*III sites in pTrc99a-cm vector. The *gill-bla* gene was custom synthesized (Genscript, NJ, USA) with *Nhe*I (5'end) and *Bgl*II (3'end) restriction enzyme sites and cloned into similar sites in pWAM1097kan_Bla to make pWAM1097kan_gill-bla plasmid.

5.4.2 Site-Directed Mutagenesis

The synonymous rare codon mutants were first designed *in silico* using a

Table 5.1. (**a**) plasmids, (**b**) strains, and (**c**) primers used in this study. The restriction enzyme sites are underlined in the primer sequences.

(a) Plasmid	**Genes**	**Copy no.**	**Resistance**	**Source**
pWAM716	hlyBD	low	Chloramphenicol	[21]
pWAM1097_ HisHlyA	6X-His tagged hlyA	high	Ampicillin	[12]
phlyM1, phlyM2, phlyEMut	HisHlyA	high	Ampicillin	Site directed mutagenesis of pWAM1097_ HisHlyA
pWAM1097 kan_Bla	bla	high	Kanamycin	this study
pblaM1, pblaM2, pblaM3	bla	high	Kanamycin	Site directed mutagenesis of pWAM1097kan_ Bla
pCMV-SPOT6_IL6	il6	high	Ampicillin	ATCC # MGC-9215
pWAM1097_I L6	il6	high	Ampicillin	this study
pWAM1097 _coIl6	coil6	high	Ampicillin	this study
pcoil6M1, pcoil6M2	coil6	high	Ampicillin	Site directed mutagenesis of pWAM1097_IL6
pET9a24a-Syn-NYESO1	nyeso1	high	Kanamycin	Gift from Ludwig Foundation for Cancer Research
pWAM1097_ ESO1	nyeso1	high	Ampicillin	this study
pesoM1, pesoM2, pesoM3	nyeso1	high	Ampicillin	Site directed mutagenesis of pWAM1097_ESO1
pWAM1097k an_GILL-Bla	gill-bla	high	Kanamycin	this study
pgillM1, pgillM2, pgillM3	gill-bla	high	Kanamycin	Site directed mutagenesis of pWAM1097_ GILL-Bla

Table 5.1 (Continued)

(a) Plasmid	Genes	Copy no.	Resistance	Source
pTrc99a-cm_ Sec-bla	bla tagged with Sec signal peptide	high	Chloramphenicol	this study
pSec-blaM1, pSec-blaM2, pSec-blaM3	bla tagged with Sec signal peptide	high	Chloramphenicol	Site directed mutagenesis of pTrc99a-cm_ Sec-bla
pTrc99a-cm_ Tat-Bla	bla tagged with Tat signal peptide (ssTorA)	high	Chloramphenicol	[22]
pTat-blaM1, pTat-blaM2, pTat-blaM3	bla tagged with Tat signal peptide	high	Chloramphenicol	Site directed mutagenesis of pTrc99a-cm_ Tat-bla
phasADE	hasADE	low	Chloramphenicol	[23]
phasM1, phasM2	hasADE	low	Chloramphenicol	Site directed mutagenesis of phasADE

(b) Strain	Plasmids	Origin
W3110	None	
MC4100	None	
CK1953	None	MC4100 *secB*::Tn5 [24]
Hly-parent	pWAM1097_HisHlyA, pWAM716	[12]
Hly-M1	phlyM1, pWAM716	W3110
Hly-M2	phlyM2, pWAM716	W3110
Hly-EMut	phlyEMut, pWAM716	W3110
Bla-W	pWAM1097kan_Bla, pWAM716	W3110
Bla-M1	pblaM1, pWAM716	W3110
Bla-M2	pblaM2, pWAM716	W3110
Bla-M3	pblaM3, pWAM716	W3110
coIL6	pWAM1097_coIL6, pWAM716	W3110
IL6	pWAM1097_IL6, pWAM716	W3110
coil6M1	pcoil6M1, pWAM716	W3110
coil6M2	pcoil6M2, pWAM716	W3110
ESO-W	pWAM1097_ESO1, pWAM716	W3110
ESO-M1	pesoM1, pWAM716	W3110
ESO-M2	pesoM2, pWAM716	W3110

Table 5.1 (continued)

(b) Strain	Plasmids	Origin
ESO-M3	pesoM3, pWAM716	W3110
GILL-W	pWAM1097kan_GILL-Bla, pWAM716	W3110
GILL-M1	pgillM1, pWAM716	W3110
GILL-M2	pgillM2, pWAM716	W3110
GILL-M3	pgillM3, pWAM716	W3110
Sec-W	pTrc99a-cm_Sec-bla	W3110
Sec-M1	pSec-blaM1	W3110
Sec-M2	pSec-blaM2	W3110
Sec-M3	pSec-blaM3	W3110
Tat-W	pTrc99a-cm_Tat-bla	W3110
Tat-M1	pTat-blaM1	W3110
Tat-M2	pTat-blaM2	W3110
Tat-M3	pTat-blaM3	W3110
Has-W	phasADE	MC4100 & CK1953
Has-M1	phasM1	MC4100 & CK1953
Has-M2	phasM2	MC4100 & CK1953

(c) Primer	Sequence
#1	5′-GCA GAA GCT TGC CAG TAC CCC CAG GAG AAG-3′
#2	5′-GAC AGA TCT CAT TTG CCG AAG AGC CCT C-3′
#3	5′-GAC AGA TCT TTA GCC TAT GGA AGT CAG G-3′
#4	5′-CTG AAT TCT TAT GCT GAT GCT GTC AAA G-3′
#5	5′-CAG TGC TAG CAC CAG TAC CCC CAG GAG AAG-3′
#6	5′-GCA GAA GCT TTT ATG CTG ATG CTG TCA AAG-3′
#7	5′-CAG TGC TAG CAG CGC CGG TTC CGC CGG GTG-3′
#8	5′-CAG TGC TAG CAC ACC CAG AAA CGC TGG TGA AAG T-3′
#9	5′-GTA CAG ATC TCC AAT GCT TAA TCA GTG AGG CAC C-3′
#10	5′-CAG TGC TAG CAA TGA GAG GAT CGC ATC ACC AT-3′
#11	5′-GTA CAG ATC TAC GGC GTT GCC CTG ATG GAG G-3′
#12	5′-CCA CAC AGA CGT CGG AAT TGC CAG CTG-3′
#13	5′-GCA GTA GTA CTT CAG AAG AAC TCG TCA AGA AGG CGA TAG-3′
#14	5′-GCG ATG GAG CTC ATG AGT ATT CAA CAT TTC CGT GT-3′
#15	5′-ATG GTG AAG CTT TTA CCA ATG CTT AAT CAG TGA GGC-3′

mathematical model of translation previously developed [25], which calculates the translation rate of the target gene expressed in *E. coli*. Primers were obtained from IDT Technologies (IO, USA). The Stratagene QuikChange II site-directed mutagenesis kit (Stratagene, CA, USA) was used per manufacturer's protocol, with a Bio-Rad *icycler*™ thermocycler (Bio-Rad, CA, USA). Confirmation of the desired changes was obtained by sequencing of purified plasmids at the Cornell Biotechnology Resource Center (Cornell University, USA).

5.4.3 Liquid blood lysis assay

The liquid blood lysis assay was adapted from several previously reported hemolysis assays [26,27]. HlyA secreting cells were grown to mid-logarithmic phase in tryptone water (1% tryptone, 0.5% NaCl) at 37°C with shaking at 250 rpm. Cultures were diluted to OD_{600} = 0.1 and grown for 1–3 hr at 37°C with shaking at 250 rpm. Cells were again diluted to OD_{600} = 0.1 and centrifuged. The supernatant was removed and serially diluted over three orders of magnitude. Sheep erythrocytes (Hardy Diagnostics) were washed at least three times in 0.9% sodium chloride by centrifugation to remove hemoglobin from lysed cells. A 4% sheep erythrocyte suspension was made in 0.9% sodium chloride. Hemolysis was monitored in 96-well plates. A reaction buffer was made consisting of 0.9% sodium chloride with 10 mM calcium chloride. Each well contained 80 µL of the diluted supernatant, 100 µL reaction buffer, and 20 µL of 4% sheep erythrocytes. Diluted supernatants were assayed in triplicate. Plates were mixed using a Versamax microplate reader (Molecular Devices) for 15 s to begin the reaction, and then incubated at 37°C. At one hour intervals, the plates were mixed for 15 s, followed by an OD_{530} measurement. Undiluted supernatant and tryptone water were used as controls. Hemolysis calculations were based on the differences between diluted supernatant samples and

controls. The predicted dilution required for 50% hemolysis was calculated by fitting a line to the slope of the lysis curve.

5.4.4 Protein fractionation and cell growth assay

Single colonies of the cells transformed with the appropriate plasmids were grown in 25 ml cultures at 37°C in Luria-Bertani (LB) supplemented with appropriate antibiotics (75 μg/mL ampicillin, 50 μg/mL kanamycin, 85 μg/mL chloramphenicol). Cells containing HasA plasmids were grown in M9 minimal media containing 0.1mg/mL thiamine and 0.4% glycerol. Aliquots were harvested at similar OD_{600} for all the samples and pelleted by centrifugation for 10 min at 4°C and 4000 × g. An aliquot of 500 μL of the supernatant fraction was transferred to another tube and mixed with two volumes of ice-cold ethanol. The tubes were left overnight at -20°C for protein precipitation and centrifuged the next day at 5000 rpm for 10 min at 4°C. The resulting protein pellets were resuspended in 20 μL of 1X phosphate buffered saline (PBS) and analyzed using 1-D SDS-PAGE and Western analysis. For the intracellular protein fractionation from whole cell pellets, the pellets were dissolved in 100 μL of SDS sample buffer, heated at 94°C for 10 min, and analyzed by SDS-PAGE and Western. Subcellular fractionation was performed using the ice-cold osmotic shock procedure [22,28]. Osmotic shockate (i.e., periplasmic fractions) was assayed for beta-lactamase activity based on nitrocefin hydrolysis in 96-well format as described elsewhere [29]. Screening of cells on solid plates was performed by adopting the method from [30]. Briefly, cells were grown for 6-7 hrs in LB medium containing 85 μg/mL chloramphenicol. 5 μL of 10X diluted cells were plated directly onto LB agar plates supplemented with 100 μg/mL ampicillin or 85 μg/mL chloramphenicol and grown overnight at room temperature.

5.4.5 Western Analysis

Supernatant fractions and intracellular protein fractions (soluble and inclusion body) were resolved by SDS–12% (w/v) PAGE using Tris–HCl and immunoblotted. Mouse anti-6XHis (1:3,000; Sigma, MO, USA) antibody was used as the primary body for the detection of HlyA and NY-ESO-1 because the two proteins contain a N-terminus 6X-His tag. Anti-HasA antibody was kindly donated by Cécile Wandersman (Institut Pasteur, France). Rabbit anti-Bla (1:3000; Millipore, MA, USA) and rabbit anti-hIL-6 (1:10,000; Sigma, MO, USA) were used for the detection of Bla and hIL-6 respectively. Secondary antibodies used were goat anti-mouse and goat anti-rabbit conjugated with alkaline phosphatase (Sigma, MO, USA) each diluted 1:30,000. Bound antibodies were detected by chemifluorescence using ECF Western kit (GE Amersham Biosciences, NJ, USA) following manufacturer's instructions and imaged using an FLA-3000 Fujifilm scanner. Quantitative analysis of the Western blots was done using ImageMaster 2D Platinum Software v5.0 (GE Amersham Biosciences)

5.4.6 hIL-6 ELISA

hIL-6 concentrations in the supernatant samples were measured by human IL-6 Quantikine ELISA kit (R&D Systems, MN, USA), per the manufacturer's protocol. It is a sandwich enzyme-linked immunosorbent assay using a mouse monoclonal capture antibody and a polyclonal antibody conjugated to horseradish peroxidase for immunodetection of bound IL-6. hIL-6 concentrations in the test samples were calculated by linear regression.

5.4.7 Co-Immunoprecipitation (CoIP)

The protocol was adapted from (ref. 31). An equivalent number of cells were harvested and the cytoplasmic fraction was isolated using the ice-cold shock procedure

[22,28]. Dynabeads® Protein A (Invitrogen, CA, USA) were mixed thoroughly by vortexing for 2-3 minutes to obtain a homogenous suspension. 100 µL of the solution containing the beads was transferred to a 2 mL tube and the beads were washed three times in 500 µl of 0.1 M Na-phosphate buffer (pH 8) and resuspended in 90 µL of the same buffer. An appropriate amount of the anti-Bla antibody (Millipore, MA, USA) was added to the beads, per manufacturer's protocol, and the antibody-bead mixture was incubated on a rolling platform for 1 hour. The tube was then placed on a magnet and the beads were washed 4 times in 500 µL of 0.1 M Na-phosphate buffer (pH 8). The cytoplasmic protein fraction was added to the beads and the mixture was incubated on the rolling platform for 2 hours at 4 °C. The beads were washed 4 times in NETN buffer (20mM Tris (pH 8), 1 mM EDTA, 900 mM NaCl) and then washed in NETN buffer containing 100 mM NaCl instead of 900 mM NaCl. The supernatant was removed and the beads were subjected to 1-D SDS-PAGE analysis by suspending the beads in 100 µL of SDS sample buffer and boiling at 95 °C for 5 mins.

5.4.8 In-gel Digestion and Mass Spectrometry

Samples were digested and characterized as described elsewhere [32]. Briefly, spots were excised using a commercially available 1.5 mm spot cutter (The Gel Company, CA, USA). Gel pieces were washed in 50% acetonitrile and digested using the ProGest automated digest station (Genomic Solutions Inc., MI, USA) and sequencing grade modified trypsin (Promega Corporation,WI, USA). Digests derived from Sypro Ruby stained gels were desalted using C18 ZipTips (Millipore, MA, USA). Digest samples were spotted onto MALDI plates using the dried droplet method with 1 µL digest and 0.5 µL 5 mg/mL α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% trifluroacetic acid. Samples were analyzed using an Applied Biosystems 4800 Proteomics Analyzer (Applied Biosystems, CA, USA) via matrix assisted laser

desorption ionization tandem time-of-flight mass spectrometry (MALDI-TOF/TOF MS). Protein identifications were based on a combination of PMF and tandem MS spectra. Spectra peaks were searched using Mascot version 2.0 [33] via the GPS Explorer software version 3.6 (Applied Biosystems, CA, USA) using NCBInr database (last modified 8/15/2006 and available at NCBI www.ncbi.nlm.nih.gov). Protein identifications with GPS Explorer confidence interval (C.I.%) scores greater than 99% were accepted. For MS/MS spectra, ion C.I.% scores greater than 99% were accepted.

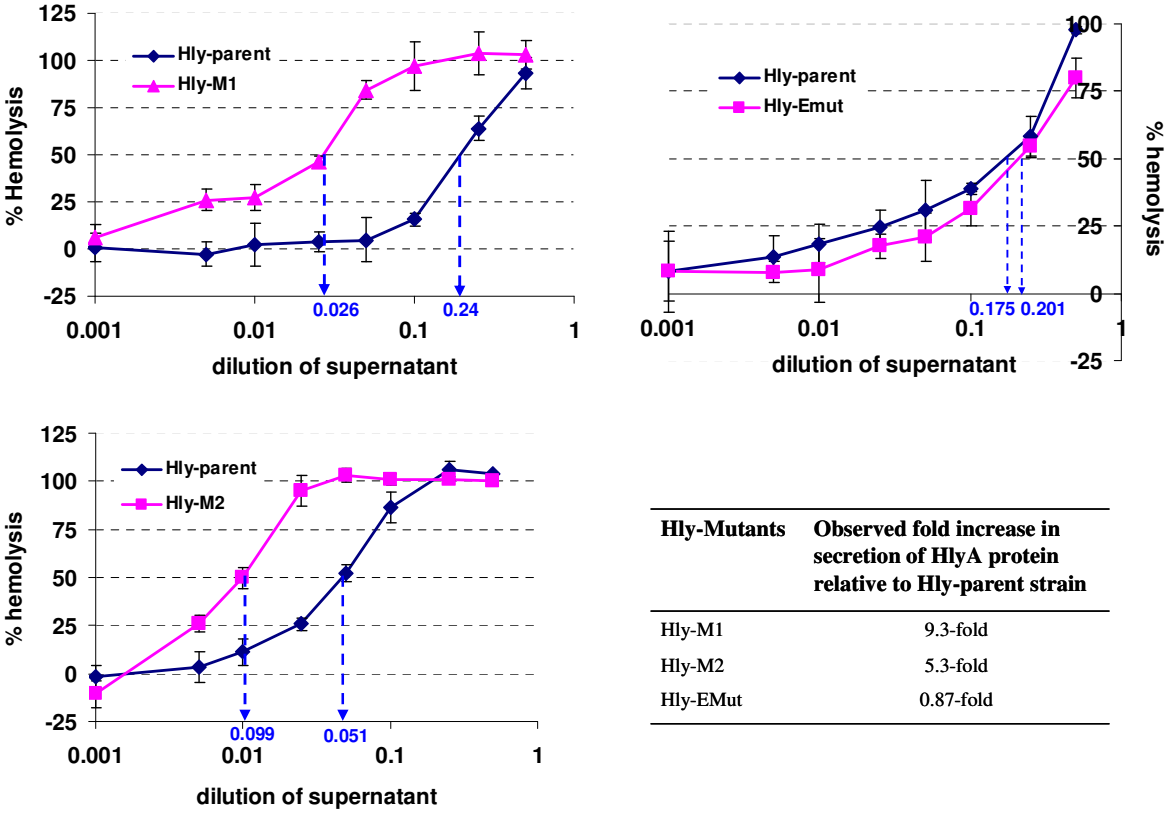
5.5 Results

Synonymous rare codon cluster can enhance secretion of recombinant protein by the Type-I pathway.

The effect of synonymous rare codon clusters on hemolysin (HlyA) secretion by the Type-I secretion pathway is tested by making substitutions at different positions. The Type-I pathway exports the polypeptides to the extracellular medium and uses a 60 amino acid C-terminal signal sequence. Previously, it has been shown that a synonymous rare codon cluster near the 5'end of the *hly*A gene results in an 8-fold increase in active HlyA detected in the culture medium [11]. Three other synonymous mutants of *hly*A gene sequence are expressed, each incorporating a rare codon cluster at different positions along the *hlyA* sequence (Table 5.2). We observe a 9.3-fold and a 5.8-fold increase in active HlyA secretion in Hly-Mut1 and Hly-Mut2 mutants relative to the Hly-parent strain respectively (Fig. 5.1) as measured by a liquid blood lysis assay [26,27]. Intracellular HlyA aggregation is decreased by nearly 50% in the Hly-Mut1 and Hly-Mut2 strains not expressing the HlyBD secretion machinery (Fig. 5.2), consistent with the predicted decrease in translation rate. No significant change in secreted active HlyA, or in intracellular HlyA in a secretion deficient strain is

Table 5.2. Sequence changes (marked in red) and predicted % decrease in translation rate of different hly-mutants

(a) Hly Mutants	Sequence		% predicted Slowdown
hlyM1	1032 CTTGGAT ACGATGGTGA CAGTTTACTT	initial sequence	39%
	CTAGGGT ACGATGGGGA CAGTCTACTA	rare codon seq.	
	344 L G Y D G D S L L	amino acid seq.	
hlyM2	1941 GGTTATC TGACCATTGA TGGCACA	initial sequence	19%
	GGTTACC TAACCATAGA TGGGACA	rare codon seq.	
	647 G Y L T I D G T	amino acid seq.	
hlyEMut	1860 AAGGTCT TTTTATCTGC CGGCTCAGCC	initial sequence	0%
	AAGGTCT TTCTATCGGC GGGCTCGGCC	rare codon seq.	
	620 K V F L S A G S A	amino acid seq.	



Hly-Mutants	Observed fold increase in secretion of HlyA protein relative to Hly-parent strain
Hly-M1	9.3-fold
Hly-M2	5.3-fold
Hly-EMut	0.87-fold

Figure 5.1. Effect of synonymous rare codon cluster on HlyA secretion by the Type-I pathway. Liquid blood assay for Hly-Mut1, Hly-Mut2, and Hly-EMut strains. Each reading performed in triplicate and error bars are the standard deviation of measurements. Fold change calculations made as follows. The supernatant dilution corresponding to 50% hemolysis is calculated (numbers in bold marked by red dashed arrows) from the lysis curve and secretion fold change is calculated by dividing this dilution. The table summarizes the change in secretion in the mutants relative to the wild-type.

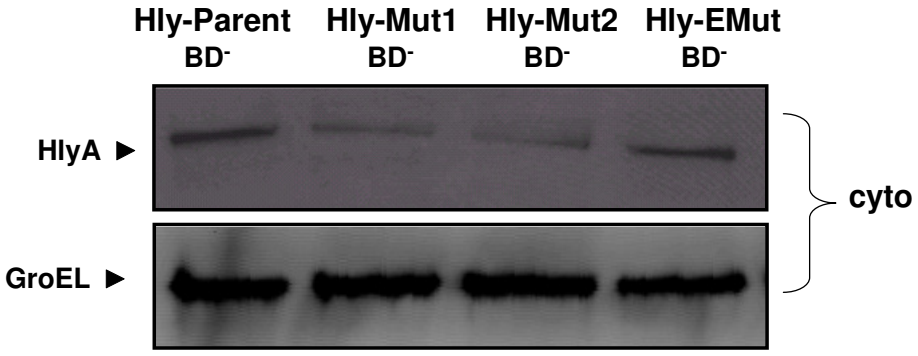


Figure 5.2 Normalized Western analysis of intracellular HlyA protein (cyto) in different hly-mutants. Secretion deficient cells (denoted by BD^-) are used for this analysis to decouple the effects of intracellular HlyA expression and secretion. An equivalent number of cells are harvested and GroEL serves as an intracellular loading control.

observed for Hly-EMut (Fig. 5.1 and 5.2), which has the same predicted translation rate relative to the parent *hlyA* gene. The presence of a synonymous rare codon cluster can also affect message stability by modulating mRNA secondary structure [34] and an analysis of the difference in the Gibbs free energy of the tested mutant *hlyA* mRNAs relative to parent mRNA indicated no significant change (Table 5.3). Comparative RT-PCR measurement of *hly*A mRNA levels (Fig. 5.3) demonstrates no significant difference either.

To test the observed effect of synonymous changes yielding enhanced secretion, beta-lactamase (Bla) secretion by the Type-I pathway was also investigated. For this experiment, the native signal sequence of Bla protein was replaced by the C-terminal Type-I signal sequence. Three synonymous mutants of the *bla* gene are expressed (Table 5.4). Quantitative western analysis reveals a 2.3-fold increase in Bla secretion in the Bla-M3 strain that contains the rare codon cluster near the 3' end of the *bla* gene and has a 21% predicted decrease in the translation rate (Fig. 5.4). The increase in extracellular Bla secretion occurs as we observed a decrease in the intracellular Bla expression in a secretion deficient strain (Fig. 5.4). As before, no significant change in the free energy of the mRNA secondary structures of *bla* mutants was observed relative to the parent *bla* mRNA (Table 5.3). These observations in HlyA and Bla are consistent with the previous finding that decreasing intracellular aggregates result in enhanced secretion [12] and suggest that the production of secreted proteins may require a balance between translation and secretion.

There are a number of proteins that demonstrate a limited capability for secretion by the Type-I pathway. Using the same approach, human interleukin-6 (hIL-6) and testis cancer antigen (NY-ESO-1) were tested. For the study of hIL-6 secretion, a codon-

Table 5.3. mRNA secondary structure analysis (change in Gibbs free energy) of different (**a**) hlyA-mutants, (**b**) bla-mutants, (**c**) il6-mutants, (**d**) nyeso1-mutants, (**e**) gill-bla mutants, and (**f**) hasA-mutants relative to the corresponding parent mRNA, as predicted by the MFOLD program [35].

(a) Hly-Mutants	**ΔΔG (kcal/mol) relative to hlyA-parent mRNA**
hlyMut1	-5.97
hlyMut2	3
hlyEMut	4.47

(b) Bla-Mutants	**ΔΔG (kcal/mol) relative to bla-parent mRNA**
blaM1	-0.6
blaM2	-1.7
blaM3	-2.7

(c) IL6-Mutants	**ΔΔG (kcal/mol) relative to coil6-parent mRNA**
Il6	-10.3
coil6M1	1.9
coil6M2	1.2

(d) NYESO1-Mutants	**ΔΔG (kcal/mol) relative to nyeso1-parent mRNA**
esoM1	3.5
esoM2	-1.3
esoM3	-2.4

(e) GILL-Bla Mutants	**ΔΔG (kcal/mol) relative to gill-parent mRNA**
gillM1	-3.4
gillM2	-1.3
gillM3	-2

(f) HasA-Mutants	**ΔΔG (kcal/mol) relative to hasA-parent mRNA**
hasM1	2.2
hasM2	-0.9

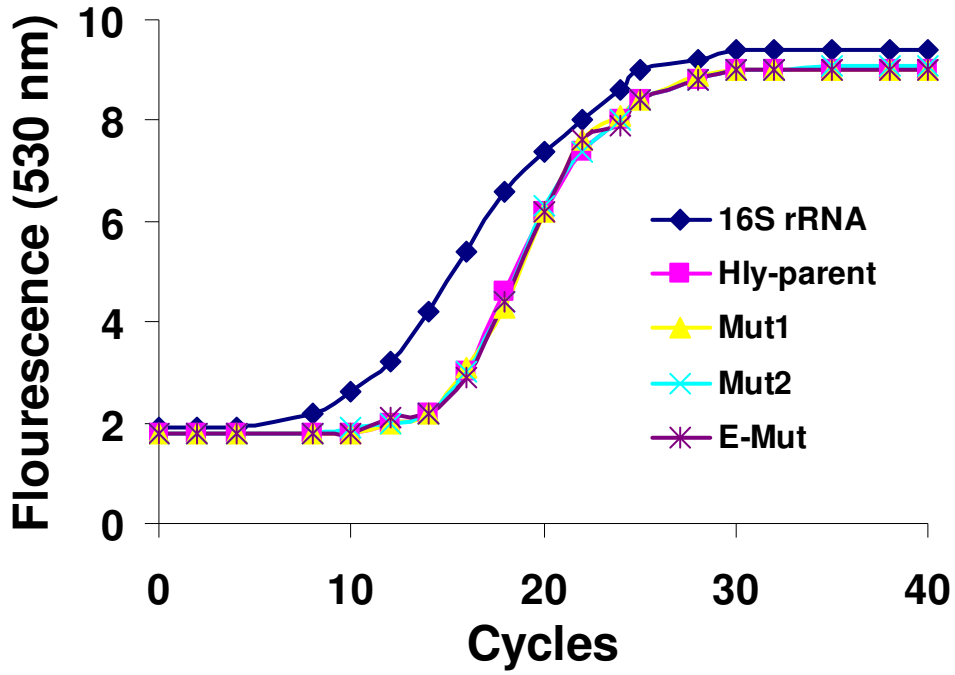


Figure 5.3. Real time RT-PCR quantifies the amount of *hly*A mRNA in different hly strains. 16S rRNA gene was used as an endogenous control.

Table 5.4. Sequence changes (marked in red) and predicted % decrease in translation rate of different bla-mutants

(b) Bla Mutants		Sequence		% predicted Slowdown
blaM1	195	CTCGGTCGCC GCATACAC	initial seq.	14%
		CTAGGGAGGA GGATACAC	rare codon seq.	
	65	L G R R I H	amino acid seq.	
blaM2	336	CTTCTGACAAC GATCGGAGGA	initial seq.	13.6%
		CTACTAACAAC GATAGGGGA	rare codon seq.	
	112	L T T I G G	amino acid seq.	
blaM3	651	CGCGGTATCA TTGCAGCA	initial seq.	21%
		AGGGGGATAA TAGCAGCA	rare codon seq.	
	217	R G I I A A	amino acid seq.	

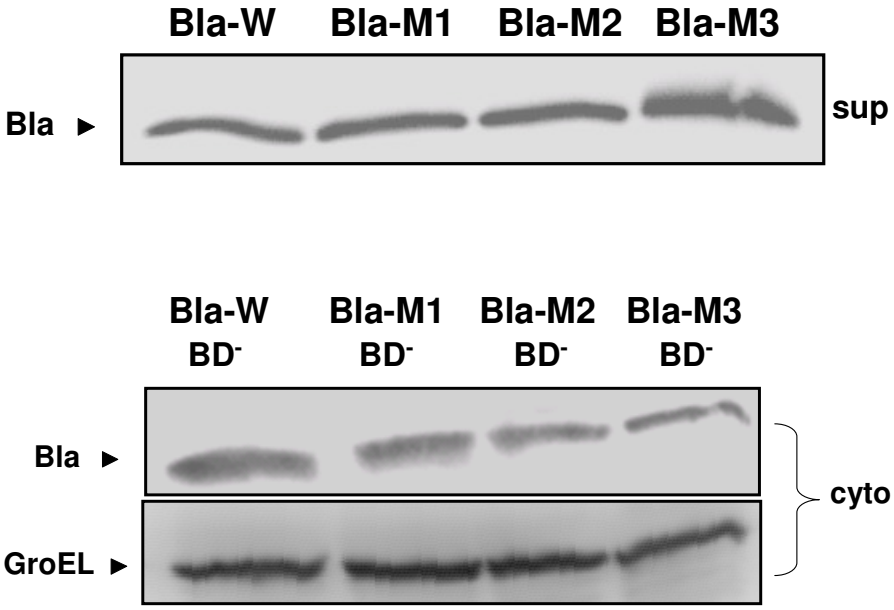


Figure 5.4. Effect of synonymous rare codon cluster on Bla secretion by the Type-I pathway Normalized Western analysis of (a) secreted (sup) beta-lactamase (Bla) protein, and (b) intracellular Bla protein (cyto) in different bla-mutants. Secretion deficient cells (denoted by BD^-) are used for this analysis to decouple the effects of intracellular Bla expression and secretion. An equivalent number of cells are harvested and GroEL serves as an intracellular loading control.

optimized interleukin-6 gene (*coil6*) with 64 codon changes relative to the native *hil6* gene was synthesized. Based on the *coil6* gene sequence, two synonymous mutants (*coil6M1* and *coil6M2*) are synthesized with a predicted decrease of 15% and 23% in the translation rate relative to the parent gene (*coIl6*), respectively (Table 5.5). Quantitative western analysis as well as ELISA, demonstrate that the IL6 strain (containing native *il6* having a 7% relative slowdown in translation rate) secretes 8-fold more IL-6 protein relative to the coIL6 strain (containing *coil6* gene) (Fig. 5.5). Moreover, a 2-fold and 2.3-fold secretion increase is observed in the coIL6-M1 and coIL6-M2 strains, respectively. NY-ESO-1 is a highly immunogenic tumor antigen and a promising vaccine candidate in cancer immunotherapy [36] and is not easily expressed in bacteria and mammalian cells [37]. Three mutants of *nyeso1* are synthesized (*esoM1*, *esoM2* and *esoM3*) with a predicted decrease of 1.9%, 20% and 15% in the translation rate relative to the *nyeso1*, respectively (Table 5.6). Quantitative western analysis shows a 3-fold and a 5-fold secretion increase in ESO-M2 and ESO-M3 strains relative to the ESO-W parent strain (Fig. 5.6). Here as well, no significant change in the free energy of the mRNA secondary structures of *il6* and *nyeso1* mutants was observed relative to the respective parent mRNA (Table 5.3). Intracellular quantitation of hIL-6 and NY-ESO-1 proteins in secretion deficient cells identified an elevated expression of the two proteins in the mutant strains relative to the parent strain (Fig. 5.5 and 5.6), a trend opposite to the results of HlyA and Bla.

Positional effects of the rare codon cluster on Bla secretion via Type-I pathway. The secretion enhancement effect of the synonymous rare codon cluster appears to be a function of the type of codon used and the position of the cluster. We investigate the positional effects of the synonymous rare codon cluster on Bla secretion by the Type-I pathway. A *de novo* gene sequence based on the *bla* backbone is designed by

Table 5.5. Sequence changes (marked in red) and predicted % decrease in translation rate of different il6-mutants

(c) IL6 Mutants	Sequence	% predicted Slowdown
il6	Human Interleukin-6 gene (64 codon changes)	7%
coil6M1	99 TACATCCTCG ACGGCATCT CAGCC initial sequence TACATCCTAG ACGGGATAT CAGCC rare codon seq. 33 Y I L D G I S A amino acid seq.	15%
coil6M2	261 GTGAAAATC ATCACTGGT CTTTTG initial sequence GTGAAAATA ATAACTGGG CTATTG rare codon seq. 87 V K I I T G L L amino acid seq.	23%

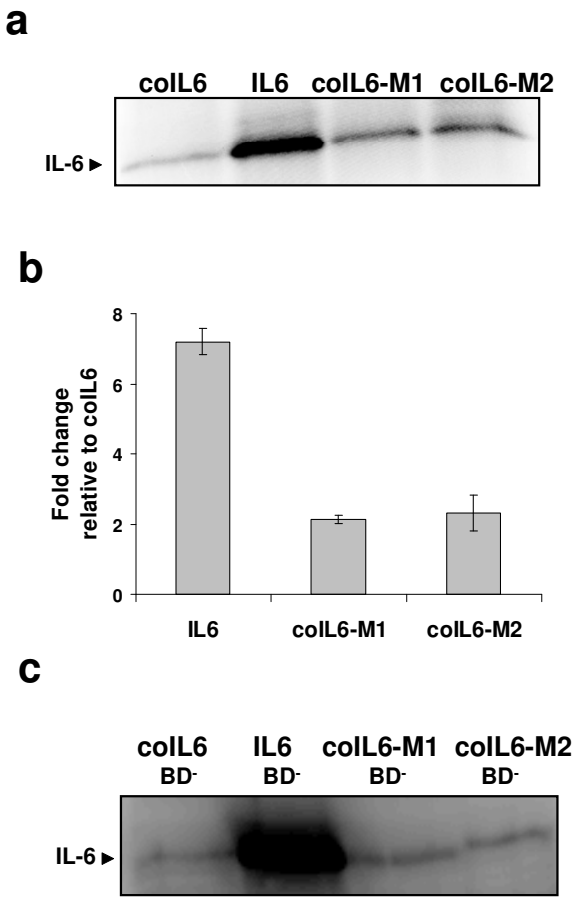


Figure 5.5. Secretion of human interleukin-6 (IL-6) protein by the Type-I pathway**.** (**a**) Normalized Western analysis of secreted IL-6. (**b**) IL-6 ELISA quantifying the amount of secreted IL-6 in different mutants relative to the coIL6 strain. ELISA results are consistent with Western analysis. (**c**) Normalized Western analysis of the intracellular IL-6 in secretion deficient cells. An equivalent number of cells are harvested.

Table 5.6. Sequence changes (marked in red) and predicted % decrease in translation rate of different nyeso1-mutants

(d) NYESO1 Mutants	Sequence			% predicted Slowdown
esoM1	87	CCAGGTGGTC CAGGTATCCC A	initial sequence	1.9%
		CCAGGTGGGC CCGGGATACC A	rare codon seq.	
	29	P G G P G I P	amino acid seq.	
esoM2	417	TCCGGCA ACATTCTGAC TATC	initial sequence	20%
		TCCGGGA ACATACTAAC TATA	rare codon seq.	
	139	S G N I L T I	amino acid seq.	
esoM3	462	CGCCAACTGC AGCTCTCCAT C	initial sequence	15%
		CGCCAACTAC AGCTATCCAT A	rare codon seq.	
	154	R Q L Q L S I	amino acid seq.	

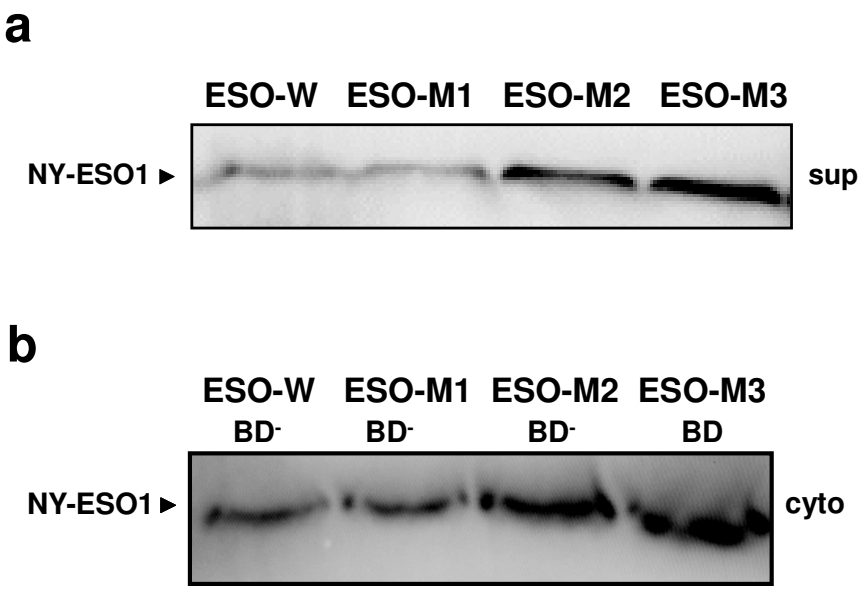


Figure 5.6. Secretion of testis cancer antigen protein (NY-ESO1) by the Type-I pathway. Normalized Western analysis of (**a**) secreted NY-ESO1, and (**b**) intracellular NYESO1 in secretion deficient cells (BD^-). An equivalent number of cells are harvested.

incorporating the codons (GGT ATC CTG CTG), encoding amino acid sequence GILL, at three different positions along the *bla* sequence (Fig. 5.7). Synonymous mutants are synthesized by replacing the new codon cluster with the 'rare' codon cluster (GGG ATA CTA CTA) at each position (Table 5.7). The introduction of these changes has no effect on the ability to quantify Bla secretion and intracellular expression (Fig. 5.8). Quantitative western analysis of the supernatant fractions indicate that GILL-M2 and GILL-M3 strains secrete approximately 1.5-fold and 2-fold more GILL-Bla protein relative to the parent strain (GILL-W), respectively (Fig. 5.8). The GILL-M2 and GILL-M3 strains have a 30 % decrease in GILL-Bla expression relative to the parent strain in secretion deficient cells (Fig. 5.8). mRNA secondary structure analysis of *gill-bla* mutants shows no significant difference relative to parent *gill-bla* mRNA (Table 5.3). Based on these observations, a rare codon cluster near the 3'end of the *bla* gene has a greater effect on the enhancement of secretion by the Type-I pathway compared to a cluster near the 5' end.

Synonymous rare codon cluster can enhance secretion via the Sec and Tat pathways and can also affect polypeptide folding.

While the effects of a synonymous rare codon cluster clearly alter the Type-I based secretion, the possibility exists that this observed phenomenon may be more general and can be tested using the general secretion pathway (Sec) and the Tat pathway in *E. coli*. The Sec and Tat pathways use an N-terminal signal peptide and secrete proteins into the periplasmic space with the note that Tat pathway is believed to secrete fully folded proteins whereas Sec does not. Three synonymous mutants (Table 5.8) were synthesized for *sec-bla* (*bla* with Sec signal sequence) and *tat-bla* (*bla* with Tat signal sequence). Cell fractionation [22,28] of different strains tracked the subcellular localization of Bla. Quantitative western analysis of the periplasmic fractions of Sec

Parent *bla* gene sequence (795 bp)

ATGCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGA
ACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTT
TTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATA
CACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGT
AAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCG
GAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAA
CCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTT
GCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGG
ATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCC
GGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTAT
CTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGA
TTAAGCATTGGTAA

Mutant *gill-bla* gene sequence (801 bp)

ATGCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGA
ACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTT
TTAAAGGTATCCTGCTGTGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGC
ATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGAC
AGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACGGTATCCTGCTGACAA
CGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGT
TGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAAC
AACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGG
AGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCT
GGAGCCGGTGAGCGTGGGTCTCGCGGTATCCTGCTGGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGT
AGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCT
CACTGATTAAGCATTGGTAA

Figure 5.7. Sequence of a *de novo* gene (*gill-bla*) to test positional dependence of synonymous rare codon cluster on Bla secretion via Type-I pathway. The synonymous codon cluster in the new gene (encoding GILL amino acid sequence) is highlighted in yellow.

Table 5.7. Sequence changes (marked in red) and predicted % decrease in translation rate of different gill-bla mutants

(e) GILL-Bla Mutants	Sequence			% predicted slowdown
gillM1	141	AAAGGTA TCCTGCTGTG T	initial sequence	5.3%
		AAAGGGA TACTACTATG T	rare codon seq.	
	47	K G I L L C	amino acid seq.	
gillM2	333	AACGGTA TCCTGCTGAC A	initial sequence	19%
		AACGGGA TACTACTAAC A	rare codon seq.	
	111	N G I L L T	amino acid seq.	
gillM3	657	CGCGGTA TCCTGCTGGC A	initial sequence	10.6%
		CGCGGGA TACTACTAGC A	rare codon seq.	
	219	R G I L L A	amino acid seq.	

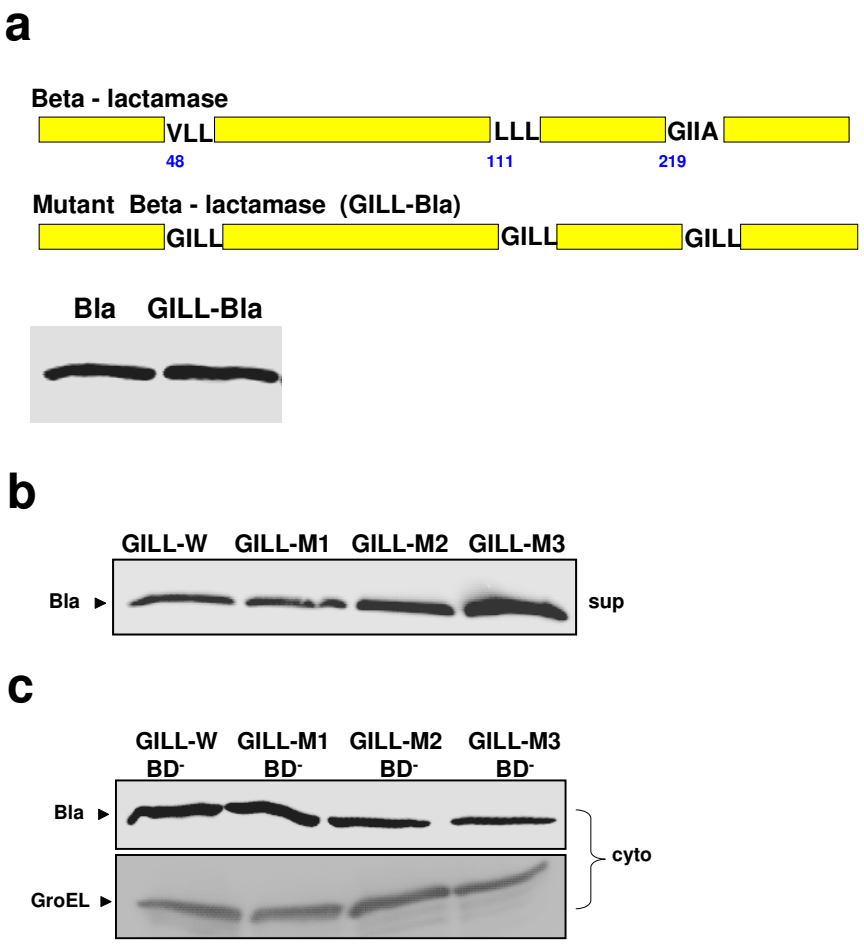


Figure 5.8. Positional effect of rare codon cluster on Bla secretion by the Type-I pathway. (a) Design of a *de novo* gene sequence (*gill*) based on *bla* gene sequence. A codon cluster (encoding amino acids GILL) is incorporated at three different positions along the gene sequence. The new protein (GILL) is detected with anti-Bla serum. Normalized Western analysis of GILL in (b) supernatant fraction (sup), and (c) intracellular fraction in secretion deficient cells (BD^-) (cyto). An equivalent number of cells were harvested and GroEL serves as an intracellular loading control.

Table 5.8. Sequence changes (marked in red) and predicted % decrease in translation rate of different sec-bla and tat-bla mutants

(b) Sec-Bla Mutants		Sequence		% predicted Slowdown
secM1	261	CTCGGTCGCC GCATACAC	initial seq.	14%
		CTAGGGAGGA GGATACAC	rare codon seq.	
	87	L G R R I H	amino acid seq.	
secM2	402	CTTCTGACAAC GATCGGAGGA	initial seq.	13.6%
		CTACTAACAAC GATAGGGGGA	rare codon seq.	
	134	L T T I G G	amino acid seq.	
secM3	717	CGCGGTATCA TTGCAGCA	initial seq.	21%
		AGGGGGATAA TAGCAGCA	rare codon seq.	
	239	R G I I A A	amino acid seq.	

(b) Tat-Bla Mutants		Sequence		% predicted Slowdown
TatM1	348	CTCGGTCGCC GCATACAC	initial seq.	14%
		CTAGGGAGGA GGATACAC	rare codon seq.	
	116	L G R R I H	amino acid seq.	
TatM2	489	CTTCTGACAAC GATCGGAGGA	initial seq.	13.6%
		CTACTAACAAC GATAGGGGGA	rare codon seq.	
	163	L T T I G G	amino acid seq.	
TatM3	804	CGCGGTATCA TTGCAGCA	initial seq.	21%
		AGGGGGATAA TAGCAGCA	rare codon seq.	
	268	R G I I A A	amino acid seq.	

strains demonstrates that the Sec-M3 strain secretes 2-fold more Bla protein than the Sec-W parent strain (Fig. 5.9). The increase in expression is consistent with an increase in Bla activity (Fig. 5.9) and the relative growth levels on solid medium containing ampicillin (Amp) (Fig. 5.9). No significant increase in Bla periplasmic secretion is observed in the Sec-M1 and Sec-M2 strains (Fig. 5.9). Quantitative western analysis of periplasmic fractions of the Tat strains reveals that Tat-M2 secretes 65% more Bla protein than the parent strain (Tat-W) (Fig. 5.10). However, a large unexpected decrease in the periplasmic secretion is observed in the Tat-M1 and Tat-M3 strains (Fig. 5.10). Bla activity data and relative growth levels of the strains on solid medium containing Amp are consistent with the Western analysis (Fig. 3e and 3f). A 1.5-fold and a 3-fold increase in cytoplasmic expression of Bla in the Tat-M2 and Tat-M3 strains relative to the Tat-W strain is also observed (Fig. 5.10). Cytoplasmic Bla activity data for Tat-M3 mutant, however, is not consistent with the Western analysis and only 20 % of the expressed Bla appears to be active (Fig. 5.10). The Tat secretion pathway has a unique ability to discriminate between properly folded and misfolded proteins *in vivo* and secrete only the folded proteins past the inner membrane [28,29]. These results suggest that the presence of the rare codon cluster at specific positions (Tat-M1 and Tat-M3 but not Tat-M2) can have an affect on Bla protein folding and subsequent recognition by the Tat secretion pathway.

Synonymous rare codon cluster affects HasA folding and its interaction with SecB chaperone.

Protein synthesis and folding in *E. coli* is co-translational [38-40] and the nucleotide sequence-dependent modulation of translation kinetics might influence nascent polypeptide folding. The HasA hemophore secretion system was used to test the effect of a rare codon cluster on polypeptide folding. HasA is a small hemoprotein (188

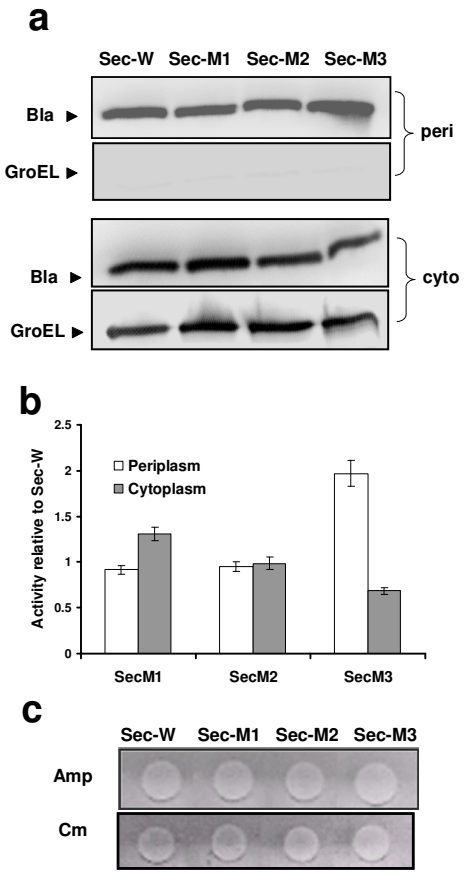


Figure 5.9. Synonymous rare codon cluster affects Bla secretion by the Sec pathway. An equivalent number of cells harvested and fractionated into cytoplasmic (cyto) and periplasmic (peri) fractions. GroEL serves as a fractionation marker. (**a**) Normalized Western analysis of the periplasmic and cytoplasmic fractions in different Sec strains. (**b**) Bla activity in periplasmic (white bars) and cytoplasmic (grey bars) fractions in Sec strains. (**c**) relative growth of Sec mutants on solid medium by spot plating 5 µL of an equivalent number of cells on LB agar supplemented with 100 µg/mL ampicillin (Amp) or 85 µg/mL chloramphenicol (Cm).

.

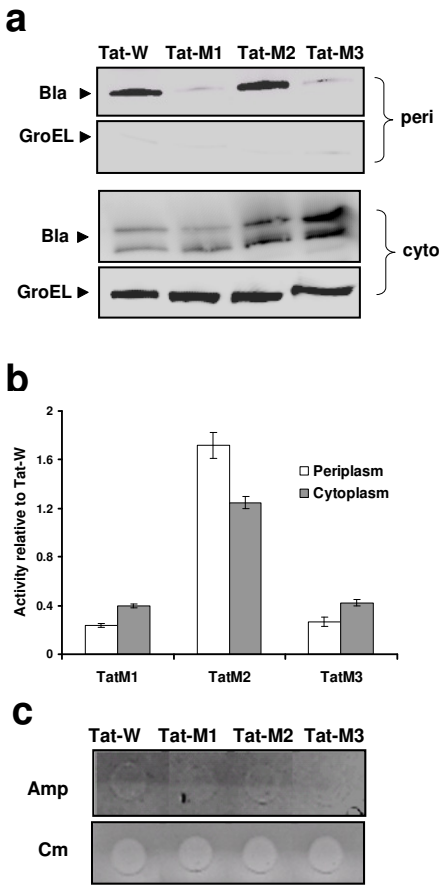


Figure 5.10. Synonymous rare codon cluster affects Bla Tatretion by the Tat pathway. An equivalent number of cells harvested and fractionated into cytoplasmic (cyto) and periplasmic (peri) fractions. GroEL serves as a fractionation marker. (**a**) Normalized Western analysis of the periplasmic and cytoplasmic fractions in different Tat strains. (**b**) Bla activity in periplasmic (white bars) and cytoplasmic (grey bars) fractions in Tat strains. (**c**) relative growth of Tat mutants on solid medium by spot plating 5 μL of an equivalent number of cells on LB agar supplemented with 100 μg/mL ampicillin (Amp) or 85 μg/mL chloramphenicol (Cm).

amino acids) of *Serratia marcescens* and is secreted by the dedicated HasA ATP binding cassette (ABC) exporter [41.]. The HasA secretion system can be successfully reconstituted in *E. coli* [41] and requires the SecB chaperone, which maintains HasA in an unfolded export-competent state [42]. HasA has a very fast folding rate and HasA that has folded in the cytoplasm can not be secreted [43,44]. If the synonymous rare codon cluster effect is general, then a synonymous rare codon cluster in *hasA* might alter HasA folding enough to permit HasA secretion in the absence of the SecB chaperone. Two synonymous mutants (*hasM1* and *hasM2*) were synthesized (Table 5.9) and have no significant change in the free energy of the mRNA secondary structures relative to the parent mRNA (Table 5.3). Quantitative western analysis was performed for HasA secretion in $secB^+$ and $secB^-$ cells are shown in Figure 5.11. No HasA secretion is observed in the parent strain (Has-W), consistent with the previous observations [42] (Fig. 5.11). However, the Has-M2 strain secretes HasA in the absence of SecB (Fig. 5.11), indicating that the rare codon cluster at this particular site alters and enhances the ability of HasA to remain in a secretion competent form. The secretion profile of the Has-M1 strain is similar to the parent strain in both genetic backgrounds (Fig. 5.11).

Rare codon cluster can affect polypeptide chaperone interactions

Changes in the translational elongation rate can affect the folding of the preceding nascent structural element or its interaction with accessory proteins, within the environment of the ribosome tunnel [45]. To test whether the presence of a synonymous rare codon cluster can affect the interaction of a polypeptide with a molecular chaperone, a co-immunoprecipitation (Co-IP) assay for Bla protein fused with the Type-I, Sec and Tat signal peptides was performed. A small number of host proteins co-immunoprecipitated with Bla (Fig. 5.12). In the Type-I and Tat secretion

Table 5.9. Sequence changes (marked in red) and predicted % decrease in translation rate of different hasA-mutants

(f) HasA Mutants		Sequence		% predicted Slowdown
hasM1	282	GGCGACGGTT TGAGCGGTGGC	initial sequence	36%
		GGGGACGGGC TAAGCGGGGGC	rare codon seq.	
	94	G D G L S G G	amino acid seq.	
hasM2	345	GGCGGCCTGA ACCTCAGC	initial sequence	40%
		GGGGGGCTAA ACCTAAGC	rare codon seq.	
	115	G G L N L S	amino acid seq.	

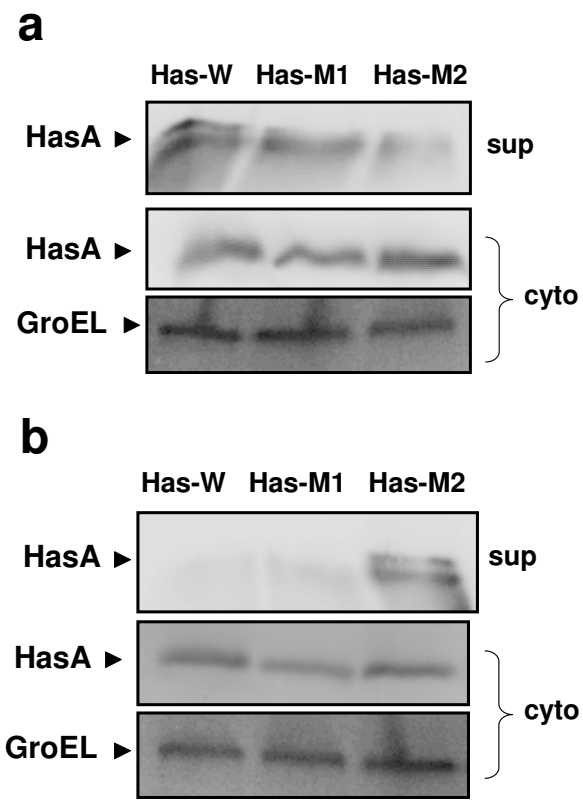


Figure 5.11. Synonymous rare codon cluster promotes HasA secretion in a SecB-independent manner. Normalized Western analysis of secreted and intracellular HasA in (**a**) MC4100 ($secB^+$), and (**b**) MC4100 ($secB^-$) strains. An equivalent number of cells were harvested and GroEL serves as an intracellular loading control. The presence of secreted HasA in the supernatant fraction of Has-M2 ($secB^-$) mutant strain indicates that the rare codon cluster in *hasM2* gene results in slower HasA folding and promote the ability to remain in a secretion-competent form.

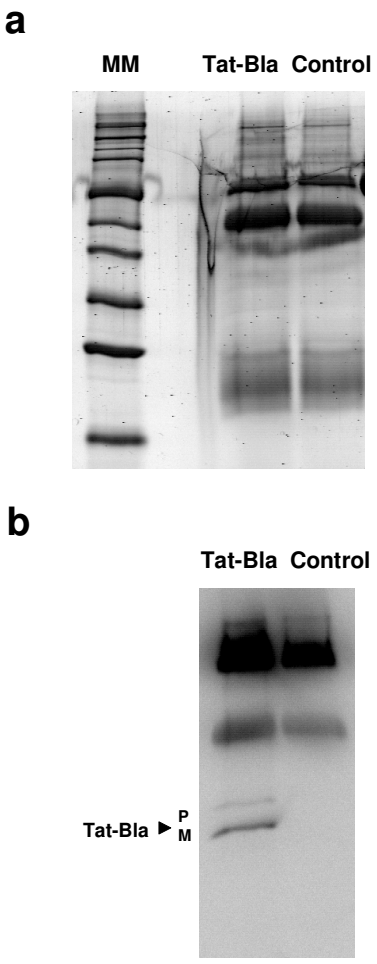


Figure 5.12. Sypro Ruby stained SDS-PAGE gel (a) and corresponding Western analysis (b) demonstrates that Bla protein expressed with the Tat signal sequence (Tat-Bla) can be immunoprecipitated from cell extracts with an anti-Bla antibody. P and M refers to precursor and mature beta-lactamase, respectively. Cells that do not express Bla protein are used as the negative control (control).

systems, Bla is found to interact with EF-Tu protein specifically identified by mass spectrometry (Fig. 5.13 and 5.14). In contrast EF-Tu did not co-immunoprecipitated with the Bla expressed using the Sec signal peptide (Fig. 5.15), demonstrating that EF-Tu may specifically interact with Bla in a secretion system dependent fashion. Previous studies reported that *E. coli* EF-Tu interacts with the hydrophobic patches of the unfolded and denatured proteins and can act as a molecular chaperone [46-49]. Here EF-Tu band intensity varied across the tested mutants. The different intensities of the EF-Tu band in different Tat samples suggests stronger association of EF-Tu in the mutants that secrete less Bla protein (Tat-M1 and Tat-M3) compared to a control (Tat-W). In contrast, the opposite trend is found for Type-I samples, where a weaker association of EF-Tu was observed for the Bla-M3 compared to the control strain (Bla-W). These data demonstrate that the synonymous rare codon cluster can alter EF-Tu interactions with the partially folded nascent Bla polypeptide, resulting in different conformations which may be either secretion competent or secretion incompetent.

5.6. Discussion

Different secretion pathways in *E. coli* were used to study the effects of synonymous substitutions on protein folding and secretion. Type-I secretion studies indicated that the synonymous rare codon cluster enhanced protein secretion, whether intracellular expression in a secretion deficient cell was decreased (HlyA, Bla) or increased (IL-6, NY-ESO-1). These results suggest that the specific effects of synonymous mutations may be protein specific and might alter the quality control mechanisms (protein aggregation, degradation) triggered by the overexpression of a recombinant protein. Increased periplasmic secretion of Bla by the Sec pathway using the synonymous codon engineering approach was also observed. Interestingly, for both the Type-I and the Sec mechanisms, the Bla mutants (Bla-M3 and Sec-M3), containing the rare codon

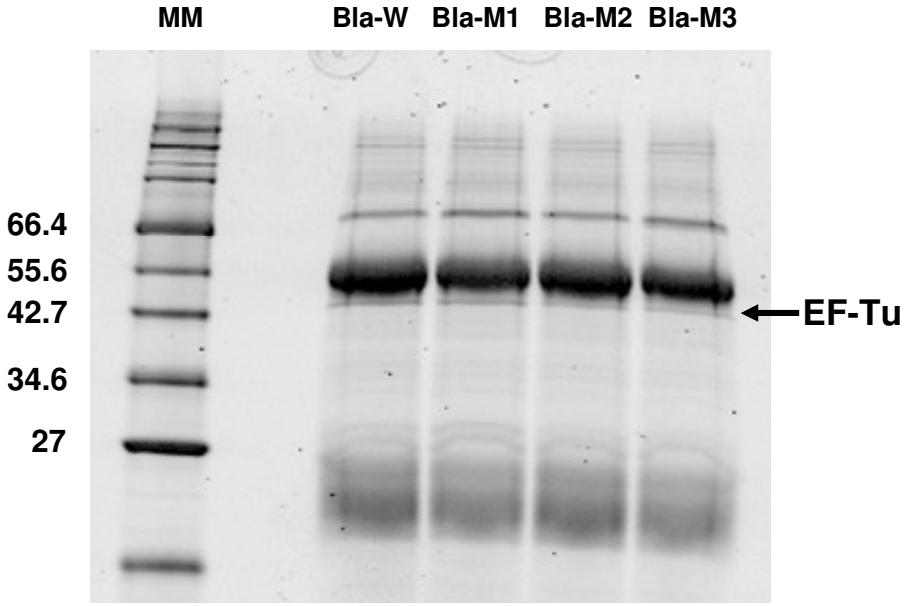


Figure 5.13. Sypro Ruby stained SDS-PAGE gel demonstrating co-immunoprecipitated proteins from different Bla strains expressed with the Type-I signal sequence. The bold arrow shows co-immunoprecipitated EF-Tu (identified by mass spectrometry). All other bands are identified as rabbit albumin, or rabbit IgG. MM refers to the molecular marker.

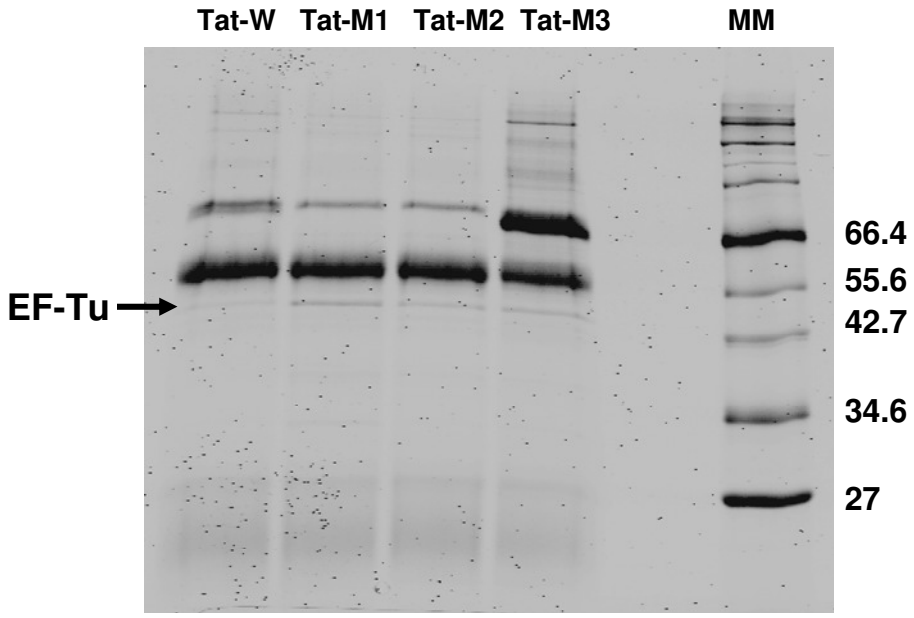


Figure 5.14. Sypro Ruby stained SDS-PAGE gel demonstrating co-immunoprecipitated proteins from different Bla strains expressed with the Tat signal sequence. The bold arrow shows co-immunoprecipitated EF-Tu (identified by mass spectrometry). All other bands are identified as rabbit albumin, or rabbit IgG. MM refers to the molecular marker.

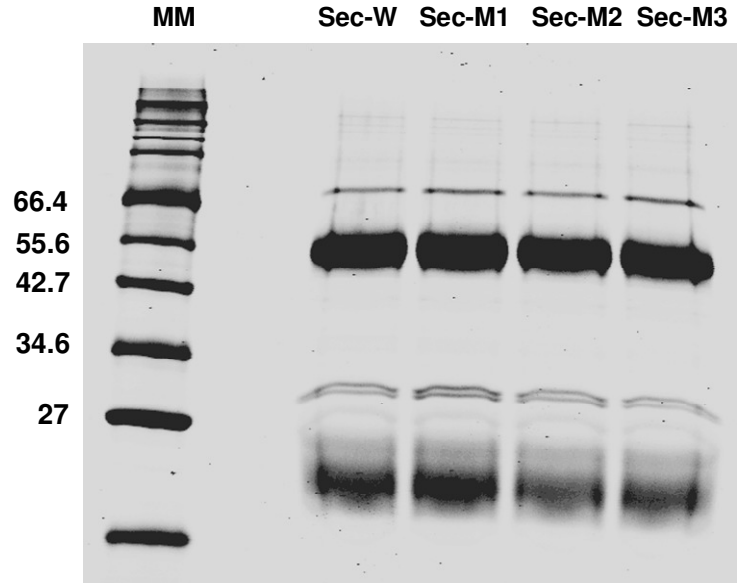


Figure 5.15. Sypro Ruby stained SDS-PAGE gel demonstrating co-immunoprecipitated proteins from different Bla mutants expressed with the Sec signal sequence. Bands are identified by mass spectrometry as rabbit albumin, rabbit IgG, or Bla. No co-immunoprecipitation of EF-Tu protein is observed.

cluster near the 3' end, secreted the most protein. It was previously observed that the C-terminal end of the Bla protein is essential to its folding and successful transport across the inner membrane [50]. While the Sec pathway translocates unfolded proteins across the cytoplasmic membrane [51], the Type-I pathway is believed to transport polypeptides in a semi-folded conformation [52]. Thus, the presence of the rare codon cluster near the 3' end may slow Bla folding, resulting in a more secretion competent polypeptide.

Synonymous rare codon clusters can significantly decrease Bla secretion in the two Tat mutants (Tat-M1 and Tat-M3) compared to the parent strain. The existence of a folding quality control mechanism is intrinsic to the Tat secretion process [22,53]. There appears to be an inconsistent relationship between the cytoplasmic Bla expression and activity in the Tat-M3 mutant compared to the parent Tat-W strain, consistent with the presence of inactive and secretion incompetent Bla. These results suggest that the presence of altered elongation rates resulting from the rare codon cluster can change the final protein conformation and activity, as has been observed in the intracellular systems not involving secretion [18-20].

The results demonstrate that the synonymous rare codon cluster can alter the interaction between the polypeptide, destined for secretion, and molecular chaperones which may aid the dynamic *in vivo* folding of the nascent polypeptide chain into different conformations. Surprisingly, the Has-M2 strain secreted HasA in a SecB-independent manner, suggesting that the synonymous rare codon cluster can alter recognition by the HasA transport machinery. Previous studies observed that reduced hydrogen bond contacts in the HasA tertiary structure can slow the HasA folding rate and reduce the SecB dependence for secretion [54]. That observation is consistent with

our finding that modulation of the elongation rates alter the folding behavior and as a result, can affect the post-translational fate of the protein. The Bla protein expressed with the Tat and Type-I secretion signal peptides interact with the EF-Tu protein whereas the Bla protein expressed with the Sec signal peptide did not. The relative EF-Tu band intensity demonstrates a stronger association of EF-Tu with the Bla polypeptide in the strains with decreased secretion for both the Type-I pathway (Bla-W and Bla-M1 relative to Bla-M3), and the Tat pathway (Tat-M1 and Ta-M3 relative to Tat-W). This analysis indicates that the interaction with EF-Tu may result in altered local secondary structures which may affect EF-Tu binding to specific sites in the polypeptide and prevent it from folding in a secretion competent conformation.

Based on these observations, there may be secretion-competent (S_C) and secretion-incompetent (S_I) pools of substrate protein inside the cell and failure to rapidly reach a secretion-competent conformation can either result in partial or complete deposition into insoluble aggregates or degradation (Fig. 5.16). The introduction of rare codons in a specific stretch of the target gene can alter the elongation rate and thus the folding of the nascent polypeptide chain enough to change the balance of these pools and therefore the secretion observed. In summary, our results offer insights into the possible mechanisms of how silent mutations affect protein folding and illustrate the utility of synonymous rare codon engineering in improving protein secretion via multiple secretion pathways for bioprocess applications.

5.7 Acknowledgements

We thank Rodney Welch for hemolysin plasmids, Matthew Delisa for pTrc99A-cm plasmid and CK1953 strain, Cecille Wandersman for phasADE plasmid and anti-HasA antibody, and Ludwig Foundation for Cancer Research for pET9a24a-Syn-NY-

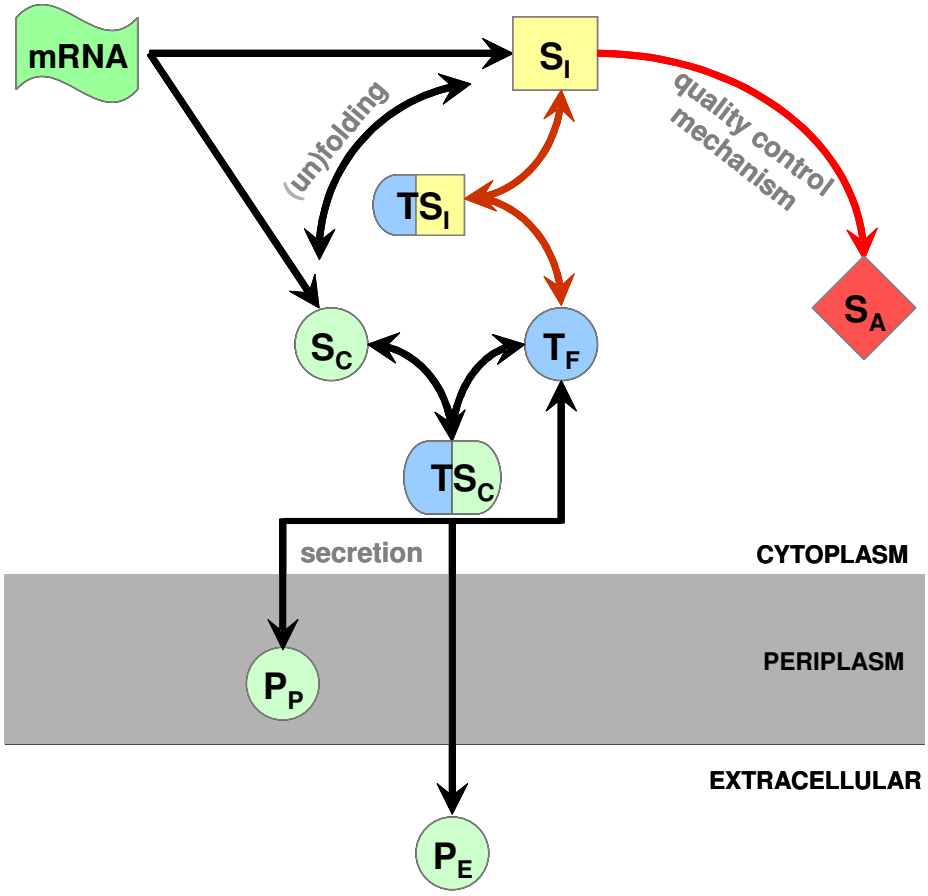


Figure 5.16. Synonymous rare codon cluster affects the balance between secretion competent (S_C) and secretion incompetent (S_I) pools of the protein, destined for secretion.

ESO-1 plasmid. We thank Leila Choe for mass spectrometric analysis and Sarah Mangan and Gautham Sreedharan for help with some of the experiments. We also thank Stephanie Hammond for helpful discussions of the manuscript. K.H.L. is supported by NYSTAR, NSF, and University of Delaware.

REFERENCES

1. Kimchi-Sarfaty C. *et al*. 2007. A‘‘silent’’ polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**: 525–528.

2. Komar AA, Lesnik T, Reiss C. 1999. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett.* **462**: 387–391.

3. Cortazzo P *et al.* 2002. Silent mutations affect in vivo protein folding in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **293**, 537-541.

4. Komar AA. 2007. Genetics. SNPs, silent but not invisible. *Science* **315**: 466-467.

5. Adzhubei AA, Adzhubei IA, Krasheninnikov IA, Neidle S. 1996. Non-random usage of 'degenerate' codons is related to protein three-dimensional structure. *FEBS Lett.* **399**: 78-82.

6. Kurland C, Gallant J. 1996. Errors of heterologous protein expression. *Curr. Opin. Biotechnol.* **7**, 489–493.

7. Lindsley D, Gallant J, Guarneros G. 2003. Ribosome bypassing elicited by tRNA depletion. *Mol. Microbiol.* **48**: 1267–1274.

8. Williams DP, Regier D, Akiyoshi D, Genbauffe F, Murphy JR. 1998. Design, synthesis and expression of a human interleukin-2 gene incorporating the codon usage bias found in highly expressed *Escherichia coli* genes. *Nucleic Acids Res*. **16**: 10453-10467.

9. Teng D, Fan Y, Yang YL, Tian ZG, Luo J, Wang JH. 2007. Codon optimization of *Bacillus licheniformis* beta-1,3-1,4-glucanase gene and its expression in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* **74**:1074-1083.

10. Kalwy S, Rance J, Young R. 2006. Toward more efficient protein expression: keep the message simple. *Mol. Biotechnol.* **34**: 151-156.

11. Disbrow GL, Sunitha I, Baker CC, Hanover J, Schlegel R. 2003. Codon optimization of the HPV-16 E5 gene enhances protein expression. *Virology* **311**: 105-114.

12. Yang B, Guo Z, Huang Y, Zhu S. 2004. Codon optimization of MTS1 and its expression in *Escherichia coli*. *Protein Expr. Purif.* **36**: 307-311.

13. Angov E, Hillier CJ, Kincaid RL, Lyon JA. 2008. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS One*. **3**: e2189.

14. Hillier CJ *et al.* 2005. Process development and analysis of liver-stage antigen 1, a preerythrocyte-stage protein-based vaccine for *Plasmodium falciparum*. *Infect. Immun.* **73**: 2109–2115.

15. Lee, PS, Lee KH. 2005. Engineering HlyA Hypersecretion in *Escherichia coli* Based on Proteomic and Microarray Analyses. *Biotechnol. Bioeng.* **89**: 195-205.

16. Gupta P, Lee KH. 2008. Silent mutations result in HlyA hypersecretion by reducing intracellular HlyA protein aggregates. *Biotechnol. Bioeng.* **101**: 967-974.

17. Etchells SA, Hartl FU. 2004. The dynamic tunnel. *Nat. Struct. Mol. Biol.* **11**: 391–392 (2004).

18. Woolhead CA, McCormick PJ, Johnson AE. 2004. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116**: 725–736.

19. Baram D, Yonath A. 2005. From peptide-bond formation to cotranslational folding: dynamic, regulatory and evolutionary aspects. *FEBS Lett.* **579**: 948–954.

20. Berisio, R. *et al.* 2003. Structural insight into the role of the ribosomal tunnel in cellular regulation. *Nat. Struct. Biol.* **10**: 366–370.

21. Felmlee T, Welch RA. 1988. Alterations of amino acid repeats in the *Escherichia coli* hemolysin affect cytolytic activity and secretion. *Proc Natl Acad Sci USA* **85**: 5269-5273.

22. DeLisa MP, Tullman D, Georgiou G. 2003. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc. Natl. Acad. Sci.* **100**: 6115–6120.

23. Ghigo JM, Le´toffe´ S, and Wandersman C. 1997. A new type of hemophore-dependent heme acquisition system of *Serratia marcescens* reconstituted in *Escherichia coli*. *J. Bacteriol.* **179**: 3572–3579.

24. Kumamoto,C.A. 1989. *Escherichia coli* SecB protein associates with exported protein precursors *in vivo*. *Proc. Natl Acad. Sci. USA* **86**: 5320–5324.

25. Shaw L, Zia R, Lee KH. 2003. Totally asymmetric exclusion process with extended objects: A model for protein synthesis. *Phys. Rev. E.* **68**: 021910.

26. Vakharia H, German GJ, Misra R. 2001. Isolation and characterization of *Escherichia coli* tolC mutants defective in secreting enzymatically active alpha-hemolysin. *J. Bacteriol.* **183**: 6908–6916.

27. Jurgens D, Ozel M, Takaisi-Kikuni NB. 2002. Production and characterization of *Escherichia coli* enterohemolysin and its effects on the structure of erythrocyte membranes. *Cell Biol. Intl.* **26**: 175-186.

28. Sargent F. *et al*. 2001. Purified components of the *Escherichia coli* Tat protein transport system form a double-layered ring structure. *Eur. J. Biochem.* **268**: 3361–3367.

29. Galarneau A, Primeau M, Trudeau LE, Michnick SW. 2002. β-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. *Nat. Biotechnol.* **20**: 619– 622.

30. Fisher A, Kim W, Delisa MP. Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. **15**: 449-458.

31. Adams PD, Ohh M. 2005. Identification of associated proteins by coimmunoprecipitation. *Nat. Methods* **2**: 475-476.

32. Finehout E, Lee KH. 2003b. A Comparison of automated in-gel digest methods for femtomole level samples. *Electrophoresis* **24**: 3508– 3516.

33. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**: 3551– 3567.

34. Nackley AG *et al.* Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* **314**, 1930–1933 (2006).

35. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31:** 3406-3415.

36. Jäger E *et. al.* 1998. Simultaneous humoral and cellular immune response against cancer–testis antigen NY-ESO-1: DeWnition of human histocompatibility leukocyte antigen (HLA)-A2-binding peptide epitopes, *J. Exp. Medi.* **187**: 265–270.

37. Piatesi A. *et al.* 2006. Directed evolution for improved secretion of cancer–testis antigen NY-ESO-1 from yeast. *Protein Expr Purif.* **48**: 232-342.

38. Kleizen B, van Vlijmen T, de Jonge HR, Braakman I. 2005. Folding of CFTR is predominantly cotranslational. *Mol. Cell* **20**: 277–287.

39. Kramer, G., Ramachandiran, V., Hardesty, B. 2001. Cotranslational folding–omnia mea mecum porto? *Int. J. Biochem. Cell Biol.* **33**: 541–553.

40. Svetlov MS, Kommer A, Kolb VA, Spirin AS. 2006. Effective cotranslational folding of firefly luciferase without chaperones of the Hsp70 family. *Protein Sci.* **15**: 242–247.

41. Ghigo JM, Le´toffe´ S, Wandersman C. 1997. A new type of hemophore-dependent heme acquisition system of *Serratia marcescens* reconstituted in *Escherichia coli*. *J. Bacteriol.* **179**: 3572–3579.

42. Delepelaire P, Wandersman C. 1998. The SecB chaperone is involved in the secretion of the *Serratia marcescens* HasA protein through an ABC transporter. *EMBO J.* **17**: 936–944.

43. Sapriel G, Wandersman C, Delepelaire P. 2003.The SecB chaperone is bifunctional in Serratia marcescens: SecB is involved in the Sec pathway and required for HasA secretion by the ABC transporter. *J. Bacteriol.* **185**: 80-88.

44. Debarbieux L, Wandersman C. 2001. Folded HasA inhibits its own secretion through its ABC exporter. *EMBO J.* **20**: 4657–4663.

45. Frydman J. 2001. Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu. Rev. Biochem*. **70**: 603-647.

46. Kudlicki W, Coffman A, Kramer G, Hardesty B. 1997. Renaturation of rhodanese by translational elongation factor (EF) Tu. Protein refolding by EF-Tu flexing. *J. Biol. Chem.* **272**: 32206-32210.

47. Richarme G. 1998. Protein-disulfide isomerase activity of elongation factor EF-Tu. *Biochem. Biophys. Res. Commun.* **252**: 156-161.

48. Malki A, Caldas T, Parmeggiani A, Kohiyama M, Richarme G. 2002. Specificity of elongation factor EF-TU for hydrophobic peptides. *Biochem. Biophys. Res. Commun.* **296**: 749-754.

49. Caldas TD, El Yaagoubi A, Richarme G. 1998. Chaperone properties of bacterial elongation factor EF-Tu. *J. Biol. Chem.* **273**: 11478-11482.

50. Koshland D, Botstein D. 1980. Secretion of beta-lactamase requires the carboxy end of the protein. *Cell* **20**: 749-760.

51. Driessen AJM. 2001. SecB, a molecular chaperone with two faces. *Trends Microbiol.* **9**: 193-196.

52. Holland IB, Schmitt L, Young J. 2005. Type 1 protein secretion in bacteria, the ABC- transporter dependent pathway. *Mol. Membr. Biol.* **22**: 29-39.

53. Sanders, C., Wethkamp, N., Lill, H. Transport of cytochrome c derivatives by the bacterial Tat protein translocation system. *Mol. Microbiol.* 41, 241–246 (2001).

54. Wolff N, Sapriel G, Bodenreider C, Chaffotte A, Delepelaire P. 2003. Antifolding activity of the SecB chaperone is essential for secretion of HasA, a quickly folding ABC pathway substrate. *J. Biol. Chem*. **278**: 38247-38253.

CHAPTER-6

WHOLE-GENOME MUTATIONAL PROFILING REVEALS A SINGLE NUCLEOTIDE POLYMORPHISM IN HYPERSECRETER *E. COLI*

6.1 Preface

DNA sequencing is a central technology in our understanding of biology and plays a significant, supporting role in drug discovery and development. The following chapter highlights the promise of short read genome sequencing technology towards rational design of interesting phenotypes for biotechnology applications and presents a single gene target in *E. coli* to enhance the extracellular secretion of recombinant proteins in *E. coli*.

6.2 Abstract

New variant organisms for biotechnology applications can be created by techniques like random mutagenesis and adaptive evolution followed by a stringent phenotypic selection. However, mutations generated in the process cannot be easily identified using traditional genetic tools. Using "next-generation" sequencing technology, a parent and a derivative hypersecreter strain (B41) of *Escherichia coli* W3110 [1] were sequenced with an average coverage of 52.8X and 55X, respectively. Mutational profiling revealed a single nucleotide polymorphism (G → T) in the B41 genome relative to the parent genome at position 1,074,787. This missense mutation results in translation termination near the N-terminal end of a transcriptional regulator protein, RutR, coded by the *ycdC* gene [2]. We verified the hypersecretion phenotype in a *ycdC*::Tn5 mutant and observed a 3.4-fold increase in active hemolysin (HlyA) secretion relative to the parent strain, consistent with the secretion increase observed in B41. mRNA expression profiling showed decreased expression of nearly all tRNA-

synthetases and some amino acid transporters in the *ycdC*::Tn5 mutant, suggesting a possible role of RutR in the translation process. This study demonstrates the potential application of high-throughput massively parallel sequencing technologies to characterize selected mutants leading to successful metabolic engineering strategies for strain improvement.

6.3 Introduction

DNA sequencing has contributed to a refined understanding of the molecular basis of many fundamental problems of biology. "Next generation" sequencing technologies and are being used extensively to understand disease association [3-5], drug resistance[6], biomarkers [7-9], and develop molecular diagnostics [10,11]. From a biotechnology standpoint, the use of these technologies to understand the genomic basis for interesting phenotypes can accelerate strain improvement by providing precise targets for metabolic engineering. Additionally, these technologies can be used in tandem with "–omics" tools to gain insights into specific gene functions. Previously, a hypersecreting *E. coli* strain (B41) was created by chemical mutagenesis of the parent strain and was found to secrete four-fold more hemolysin (HlyA) protein relative to the parent strain via the Type-I secretion pathway [2]. The genomic DNA mutations generated in the process, however could not be identified because of technological limitation at that time. In this study, we used short read sequencing technology to reveal a single nucleotide polymorphism (SNP) in the B41 genome relative to the parent genome. mRNA expression profiling of the *ycdC*::Tn5 mutant and the parent strain is performed to understand the basis of the hypersecretion phenotype in the mutant strain. This study presents a single gene target in *E. coli* to enhance the extracellular secretion of recombinant proteins and illustrates the promise of genome sequencing towards rational design of interesting phenotypes for

biotechnology applications.

6.4 Materials and Methods

6.4.1 Plasmids and strains

All plasmids and strains used in this study are shown in Table 6.1. Plasmid pWAM1097kan_Bla was created as follows. First *bla* gene was PCR amplified with primers (5'- GCA GAA GCT TGC ACC CAG AAA CGC TGG TGA AAG T -3' and 5'- GAC AGA TCT CCA ATG CTT AAT CAG TGA GGC -3') incorporating *Hind*III restriction site at the 5'end and *Bgl*II restriction site at the 3' end and HlyA secretion signal sequence (180 bp) was PCR amplified with primers (5'- GAC AGA TCT TTA GCC TAT GGA AGT CAG G -3' and 5'- CTG AAT TCT TAT GCT GAT GCT GTC AAA G -3') incorporating *Bgl*II restriction site at the 5'end and *EcoR*I restriction site at the 3' end. These fragments were ligated and then cloned into *Hind*III and *EcoR*I sites of pEGFP vector (Clontech, CA, USA). The fusion gene sequences were then PCR amplified with primers (5'- CAG TGC TAG CAC ACC CAG AAA CGC TGG TGA AAG T -3' and 5'- GCA GAA GCT TTT ATG CTG ATG CTG TCA AAG -3') incorporating *Nhe*I restriction site at the 5'end and *Hind*III restriction site at the 3' end and cloned into the same sites of pWAM1097_HisHlyA to make pWAM1097_Bla. Finally, kanamycin (*kan*) resistant marker gene was PCR amplified with primers (5'- CCA CAC AGA CGT CGG AAT TGC CAG CTG -3' and 5'- GCA GTA GTA CTT CAG AAG AAC TCG TCA AGA AGG CGA TAG -3') incorporating *Aat*II (5'end) and *Sca*I (3'end) restriction enzyme sites and then cloned into similar sites in pWAM1097_Bla to create pWAM1097kan_Bla plasmid. The plasmids were then transformed with or without pWAM716 (encoding Type-I secretion machinery) in FB22271 (MG1655*ycdC*::Tn5) and MG1655 cells to create appropriate strains.

Table 6.1. Table of plasmids and strains used in this study; (A) shows plasmids used in the study and their sources (B) shows strains used and their origin.

(A) Plasmid	**Genes**	**Copy no.**	**Resistance**	**Source**
pWAM716	hlyBD	low	Chloramphenicol	[12]
pWAM1097	hlyCA	high	Ampicillin	[13]
pWAM1097_HisHlyA	hlyC, 6X-His tagged hlyA	high	Ampicillin	[14]
pWAM1097kan_Bla	Bla	high	Kanamycin	This study

(B) Strain	**Plasmid**	**Origin**
W3110	None	
MG1655	None	(*E. coli* K-12 MG1655 Genome Initiative)
FB22271	None	MG1655 *ycdc*::Tn5 *E. coli* K-12 MG1655 Genome Initiative)
Hly-parent	pWAM1097, pWAM716	W3110-based [1]
B41	pWAM1097, pWAM716	W3110-based [1]
Hly-ycdC$^+$	pWAM1097_HisHlyA, pWAM716	MG1655-based (this study)
Hly-ycdC$^-$	pWAM1097_HisHlyA, pWAM716	FB22271-based (this study)
Hly-ycdC$^+$ (BD$^-$)	pWAM1097_HisHlyA	MG1655-based (this study)
Hly-ycdC$^-$ (BD$^-$)	pWAM1097_HisHlyA	FB22271-based (this study)
Bla-ycdC$^+$	pWAM1097kan_Bla, pWAM716	MG1655-based (this study)
Bla-ycdC$^-$	pWAM1097kan_Bla, pWAM716	FB22271-based (this study)
Bla-ycdC$^+$ (BD$^-$)	pWAM1097kan_Bla	MG1655-based (this study)
Bla-ycdC$^-$ (BD$^-$)	pWAM1097kan_Bla	FB22271-based (this study)

6.4.2 Genome sequencing and analysis

Genomic DNA of hly-parent and B41 strains was isolated using GenElute Bacterial Genomic DNA Kit (Sigma, MO, USA), according to manufacturer's instructions and the samples were sequenced in duplicates. Genome sequencing was performed and analyzed with a Illumina Genome Analyzer^{TM} (Illumina, CA, USA) at Cornell University Life Sciences Core Laboratories Center. Briefly, 5 μg of genomic DNA was fragmented below 800 bp using a nebulizer and end-repaired with T4 DNA polymerase. A single dA was added to the ends using Klenow fragment and dATP. Fragments were then ligated with adaptors provided by the manufacturer. Adaptor ligated fragments were separated from unligated adaptors by running an agarose gel and cutting a band corresponding to 150–200 bp and purified using a spin column. The fragment library containing adaptors was subjected to 10 rounds of PCR using primers supplied by Illumina. This amplified library was then loaded onto the cluster generation station for single molecule bridge amplification on slides containing attached primers. The slide with amplified clusters was then subjected to step-wise sequencing using four-color labeled nucleotides on the Illumina Genome Analyzer II for 36 cycles. A total of 15,638,760 read sequences were obtained after quality filtering.W3110 genome sequence (Genbank accession number AP009048) was used as the reference genome to align the sequences of the two genomes. Alignment was accomplished using the Eland program (Illumina, CA, USA) and up to two mismatches were allowed per read. Reads that did not align to the genome uniquely were discarded. An SNP detection program was written in Java (Sun Microsystems, CA, USA). The numbers of correctly aligned and mismatched base pairs at each position were used to generate a mismatch percentage. The difference in mismatch percentages at each position in the two alignments was used to detect SNPs.

6.4.3 Quantitative real time reverse transcription polymerase chain reaction (qRT-PCR)

*hly*A mRNA level was evaluated by comparative real-time RT-PCR utilizing TaqMan® One-Step RT-PCR Master Mix Reagents Kit and ABI 7900HT Sequence Detection System (Applied Biosystems, CA, USA). Total RNA was extracted from the cells using Qiagen RNeasy kit (Qiagen, CA, USA), as per the manufacturer's protocol and 10 ng of total RNA was used for each reaction. 16S rRNA gene was used as an internal control. The *hly*A specific forward primer 5'-GGT ATT CGG CAC AGC AGA GAA -3' and the reverse primer 5'-GTC TAA TTG TGG TGC AAA GAT AGT CAC T -3' and 16S rRNA specific forward primer 5'-CCA GCA GCC GCG GTA AT -3'and the reverse primer 5'-TGC GCT TTA CGC CCA GTA AT -3' were used in these studies. The TaqMan® Probe with 6-carboxyfluorescein (6-FAM) as the reporter dye and tetramethyl-6-carboxyrhodamine (TAMRA) as the quencher was used and the 16S rRNA probe sequence was 5'-CCG ATT AAC GCT TGC ACC CTC CG -3' and *hly*A probe sequence was 5'-CTC ATT GGC CTC ACC GAA CGG G -3'. The threshold cycle for each amplification curve was calculated using SDS software, version 2.1 (Applied Biosystems, CA, USA).

6.4.4 Vancomycin assay

This protocol is adapted from a published study [15]. The cells expressing the Type-I secretion system were grown in 25 mL cultures at 37°C in Luria- Bertani (LB) supplemented with ampicillin (75 μg/mL) and chloramphenicol (85 μg/mL) until they reached OD_{600} = 1. Thereafter, aliquots of cells were taken and incubated at 37°C for 30 min with mild shaking (200-250 rpm) in the presence of different concentrations of vancomycin antibiotic. After vancomycin treatment, cells were plated on LB-agar plates in the absence of vancomycin and supplemented with ampicillin (75 μg/mL)

and chloramphenicol (85 μg/mL). Survival was reported as a percentage of the number of colonies formed by control samples previously incubated in the absence of vancomycin.

6.4.5 Liquid blood lysis assay

The liquid blood lysis assay was adapted from previously reported hemolysis assay protocols [16,17]. HlyA secreting cells were grown to mid-logarithmic phase in tryptone water (1% tryptone, 0.5% NaCl) at 37°C with shaking at 250 rpm. Cultures were diluted to OD_{600} = 0.1 and grown for 1–3 hr at 37°C with shaking at 250 rpm. Cells were again diluted to OD_{600} = 0.1 and centrifuged. The supernatant was removed and serially diluted over three orders of magnitude. Sheep erythrocytes (Hardy Diagnostics, CA, USA) were washed at least three times in 0.9% sodium chloride by centrifugation to remove hemoglobin from lysed cells. A 4% sheep erythrocyte suspension was made in 0.9% sodium chloride. Hemolysis was monitored in 96-well plates. A reaction buffer was made consisting of 0.9% sodium chloride with 10 mM calcium chloride. Each well contained 80 μL of the diluted supernatant, 100 μL reaction buffer, and 20 μL of 4% sheep erythrocytes. Diluted supernatants were assayed in triplicate. Plates were mixed using the Versamax microplate reader (Molecular Devices) for 15 s to begin the reaction, and then incubated at 37°C. At one hour intervals, the plates were mixed for 15 s, followed by an OD_{530} measurement. Undiluted supernatant and tryptone water were used as controls. Hemolysis calculations were based on the differences between diluted supernatant samples and controls. The predicted dilution required for 50% hemolysis was calculated by fitting a line to the slope of the lysis curve.

6.4.6 Protein fractionation

Single colonies of the cells transformed with the appropriate plasmids were grown in 25 ml cultures at 37°C in Luria- Bertani (LB) supplemented with appropriate antibiotics (ampicillin (75 μg/mL), kanamycin (30 μg/mL), chloramphenicol (85 μg/mL)). Aliquots of cells were harvested at similar OD_{600} for all the samples and pelleted by centrifugation for 10 min at 4°C and 4000 × g. 500 μL of the supernatant fraction was transferred to another eppendorf tube and mixed with two volumes of ice-cold ethanol. The tubes were left overnight at -20°C for protein precipitation and centrifuged next day at 5000 rpm for 10 min at 4°C. The resulting protein pellets were resuspended in 20 μL of 1X phosphate buffered saline (PBS) and analyzed using 1-D SDS-PAGE and western blotting. For intracellular protein fractionation from whole cell pellets, the pellets were dissolved in 100 μL of SDS sample buffer, heated at 94°C for 10 min, and analyzed using SDS-PAGE and western blotting.

6.4.7 Western Analysis

Supernatant fractions and intracellular protein fractions (soluble and inclusion body) were resolved by SDS-PAGE (12% w/v) using Tris–HCl and immunoblotted. Mouse anti-6X-His (Sigma, MO, USA; 1:3000) and rabbit anti-Bla (Millipore, MA, USA; 1:3000) were used as primary antibodies and alkaline phosphatase conjugated goat anti-mouse IgG antibody (Sigma, MO, USA; 1:30,000) and anti-rabbit IgG antibody (Sigma, MO, USA; 1:30,000) were used as secondary antibodies for the detection of respective proteins. Bound antibodies were detected using enhanced chemifluorescence (ECF) substrate (GE Amersham Biosciences, NJ, USA) following the manufacturer's instructions and imaged using a FLA-3000 Fujifilm scanner. Quantitative analysis of the western blots was done using ImageMaster 2D Platinum Software v5.0 (GE Amersham Biosciences, NJ, USA).

6.4.8 mRNA Isolation and GeneChip Analysis

Three biological replicate cultures for each strain were harvested at a mid-exponential phase and were resuspended in two volumes of RNAprotect Bacterial Reagent (Qiagen, CA, USA). Total RNA was extracted from the cells using Qiagen RNeasy kit (Qiagen, CA, USA), as per the manufacturer's protocol. RNase-Free DNase kit (Qiagen, CA, USA) was also used to minimize the genomic DNA contamination. RNA samples were processed according to the Affymetrix prokaryotic sample and array processing protocol. Briefly, cDNA fragments were generated from RNA samples via reverse transcription using random hexamer primers. The resulting cDNA was then fluorescently labeled and hybridized to GeneChip® *E. coli* Genome 2.0 Array (Affymetrix, CA, USA) and the arrays were scanned at the Cornell University Life Sciences Core Laboratories Center. Probe intensity data from scanned chip images were normalized using a RMA procedure and analyzed using GeneSpring™ software (Agilent, CA, USA). The fold change in gene expression was calculated by comparing mRNA expression values across the two strains. Unpaired t-test with $P \leq 0.1$ was used to identify the genes with statistically significant expression change.

6.5 Results and Discussions

B41 and Hly-parent genomes were sequenced using the Illumina® genome analyzer. The sequence fragments were aligned to the W3110 *E. coli* genome (Genbank accession number AP009048) using the Eland program (Illumina), which allows up to two base pair (bp) mismatches between each 32 bp sequence fragment and the W3110 genome. The average depth of sequence coverage for the Hly-parent and B41 genomes was found to be 52.8X and 55.0X, respectively (Fig. 6.1). For each alignment, 97.5% of the published W3110 genome was covered by 8 sequence fragments or more. A base pair caller program was developed in Java™ (Sun Microsystems, CA, USA) to

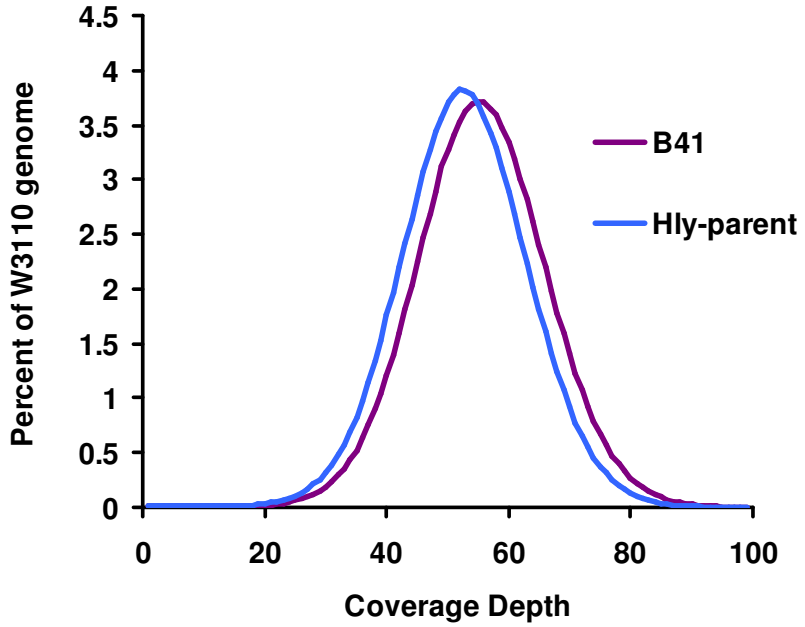


Figure 6.1. Genome coverage of W3110 *E. coli* from Illumina® sequence data. The x-axis denotes the sequence coverage depth and the y-axis shows the percentage of *E. coli* genome covered by each of the two alignments. Duplicate samples of Hly parent and B41 genomes were sequenced and analyzed.

detect single nucleotide polymorphisms. The percentage of mismatched base pairs was recorded at each position in the W3110 genome. A log plot histogram detailing the frequency of mismatch percentages throughout the genome is shown in Figure 6.2. 99.99% of the covered genome had a mismatch percentage of 20% or less. To detect mutations, the percent of mismatches at each position in the genome from Hly-parent and B41 strains were compared and the ten positions with the greatest difference were identified (Fig. 6.3). Positions with a large difference in the mismatch percentage represent probable polymorphisms between the B41 mutant and Hly-parent. The single nucleotide polymorphism (SNP) detected at position 1,074,787 has 100% difference in the mismatch rate between the B41 and Hly-parent alignments and was present in all 49 sequence reads from both B41 samples and none of the 58 sequence reads from the two Hly-parent samples (Fig.6.4). All other positions had a mismatch percentage difference of 30% or less (Fig. 6.3). The SNP (at position 1,074,787) results in a guanine to thymine mutation in the *ycdC* gene (at position 124 bp relative to the translation start site), which prematurely terminates the translation of RutR protein, a recently discovered transcriptional repressor of the pyrimidine utilization pathway (rut pathway) [2].

To understand how the absence of RutR increases active HlyA secretion in the B41 strain, three possible contributing factors were considered: higher expression of *hlyA* mRNA, higher expression of Type-I secretion machinery, and secretion of accessory proteins that may affect HlyA activity and stability. No significant difference in *hlyA* mRNA levels in the Hly-parent and B41 strains was observed by comparative real time RT-PCR (Fig. 6.5). To test for higher expression of transport machinery in the B41 strain, we relied on the observation that cells that express Type-I secretion machinery are hypersensitive to vancomycin, an antibiotic that is relatively inactive

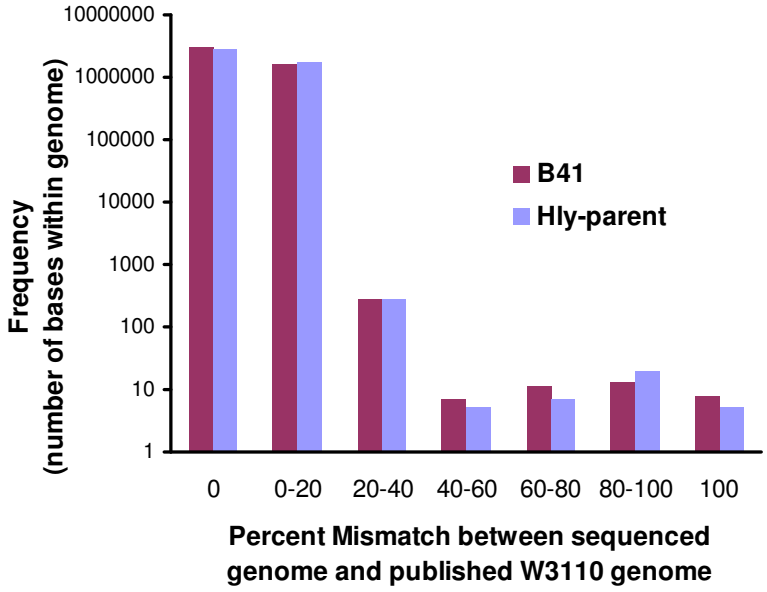


Figure 6.2. Distribution of mismatch percentages across the Hly parent and B41 genomes in log scale. The x axis shows percentages of aligned base pairs that do not match W3110 genome. Aligned fragments covered 97.3% of the W3110 genome. 99.99% of the covered portion of the genome had a mismatch percentage under 20%.

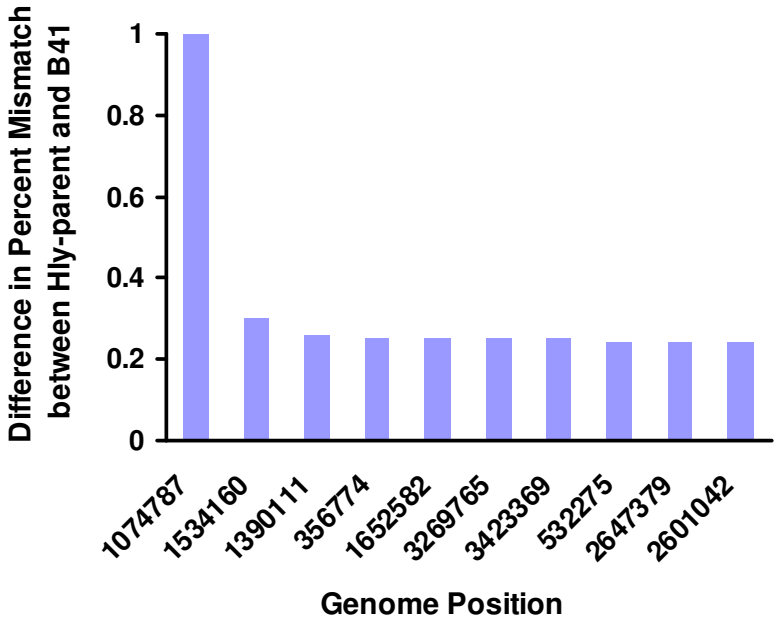


Figure 6.3. Percent differences of mismatched bases between Hly parent and B41 genomes. The positions with the ten largest values are displayed. A difference of 1 occurs when all the base pairs at a certain position in B41 genome do not match and all the base pairs at a certain position in Hly-parent genome match the W3110 genome.

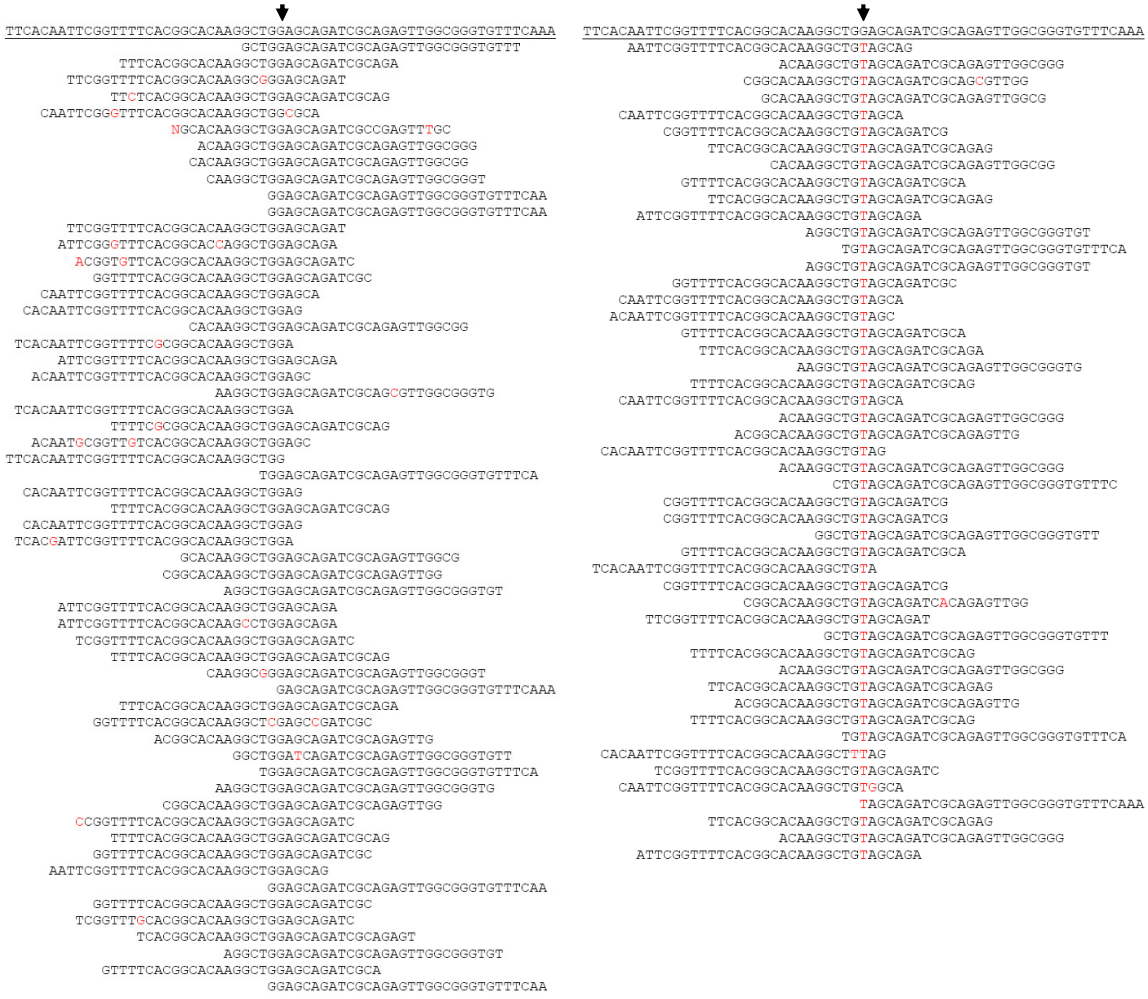


Figure 6.4. Sequence data from Hly parent and B41 genomes illustrating the presence of single nucleotide polymorphism in B41 genome. The mutation (G→T) was present in all 49 sequence reads from both B41 samples and none of the 58 sequence reads from the two hly-P samples. The underlined set of nucleotides is the reference W3110 genome and the arrow indicates the position of the SNP in the B41 mutant relative to the Hly-parent.

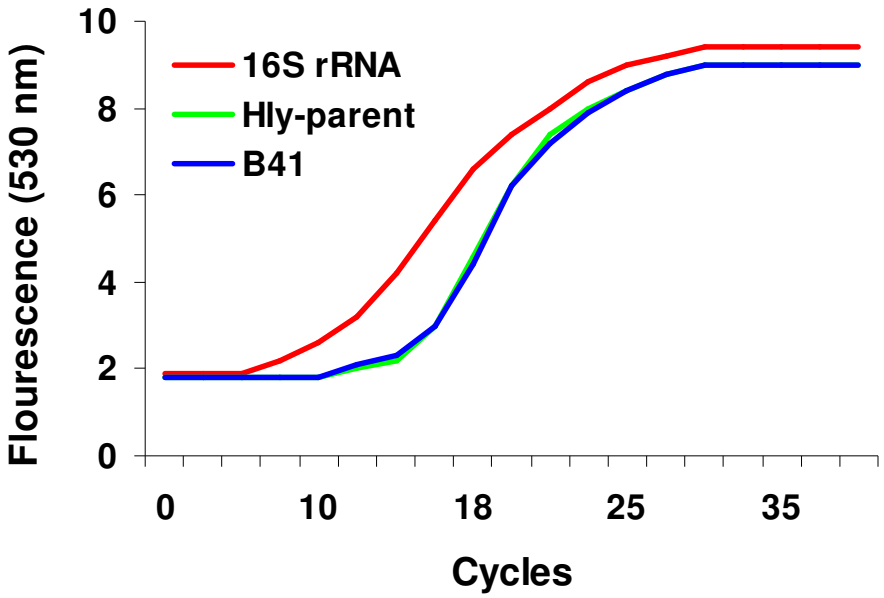


Figure 6.5. Real time RT-PCR to quantify the amount of *hly*A mRNA in Hly parent and B41 strains. 16S rRNA gene was used as an endogenous control. The fold change in *hly*A mRNA expression in the two strains was found out to be 0.98.

against gram-negative bacteria [18]. Increased vancomycin sensitivity is consistent with elevated expression levels of functional Type-I transport machinery [15]. No significant difference in cell viability was observed between the Hly-parent and B41 strains (Fig. 6.6), consistent with no significant difference in the number of Type-I transporters in the two strains. Finally, extracellular proteomes of the Hly-parent and B41 strains were compared by profiling their supernatant fractions to identify changes in the secreted proteins (Fig. 6.7). The results are qualitatively similar, suggesting that the two strains have similar extracellular proteome profiles.

The absence of functional RutR might also affect the relative HlyA protein expression in the two strains. To measure the intracellular HlyA expression, a 6X-His tag was cloned at the N-terminal end of HlyA protein. The activity level of 6X-His tagged HlyA was found to be similar to HlyA (Fig. 6.8). Secretion and intracellular HlyA quantitation studies were then performed in a ycdC negative background. Appropriate Hly-plasmids (pWAM1097_HisHlyA, pWAM716) were transformed in FB22271 (MG1655 based strain containing a *ycdC*::Tn5 lesion) and parent MG1655 strains to construct Hly-ycdC$^-$ and Hly-ycdC$^+$ strains respectively. A 3.4-fold increase in HlyA secretion was observed in the Hly-ycdC$^-$ strain relative to the Hly-ycdC$^+$ strain by western blotting (Fig. 6.9). The increase in active HlyA secretion was also confirmed by a liquid blood lysis assay (Fig. 6.9) and is consistent with the increased secretion observed in the B41 strain [1]. Intracellular HlyA quantitation was performed in secretion deficient ycdC$^-$ and ycdC$^+$ strains (denoted by Hly-ycdC$^-$ (BD$^-$) and Hly-ycdC$^+$ (BD$^-$)) to decouple the effects of secretion and HlyA protein production. Western analysis indicated a 23% decrease in HlyA protein expression in the Hly-ycdC$^-$ (BD$^-$) relative to Hly-ycdC$^+$ (BD$^-$) strain (Fig. 6.10). We also examined the secretion of beta-lactamase (Bla via the Type-I secretion pathway in a ycdC negative

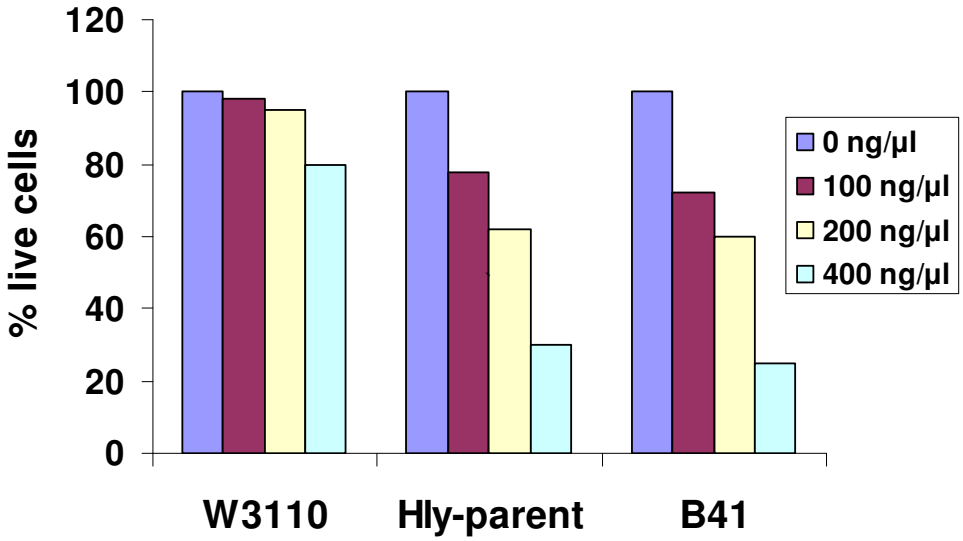


Figure 6.6. Cell viability in the presence of different concentrations of vancomycin antibiotic test whether enhanced secretion phenotype is a result of increased expression of secretion machinery. The colored bars indicate percentage of viable cells of Hly-parent and B41 strains at different concentration of vancomycin relative to the cells grown in the absence of vancomycin. W3110 is used as a control strain.

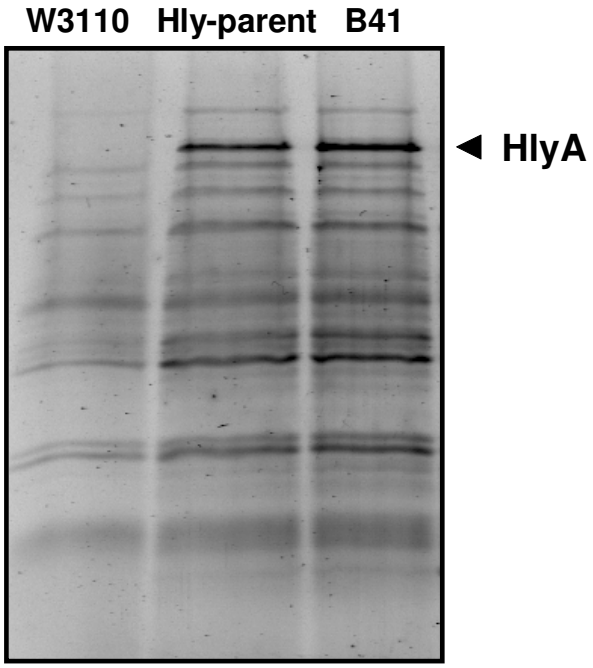


Figure 6.7. Extracellular proteome profile of Hly-parent and B41 strains using 1-D SDS-PAGE. The supernatant fractions were collected at similar cell optical density (OD_{600}). HlyA protein band (~110 KDa) was confirmed by mass spectrometry analysis.

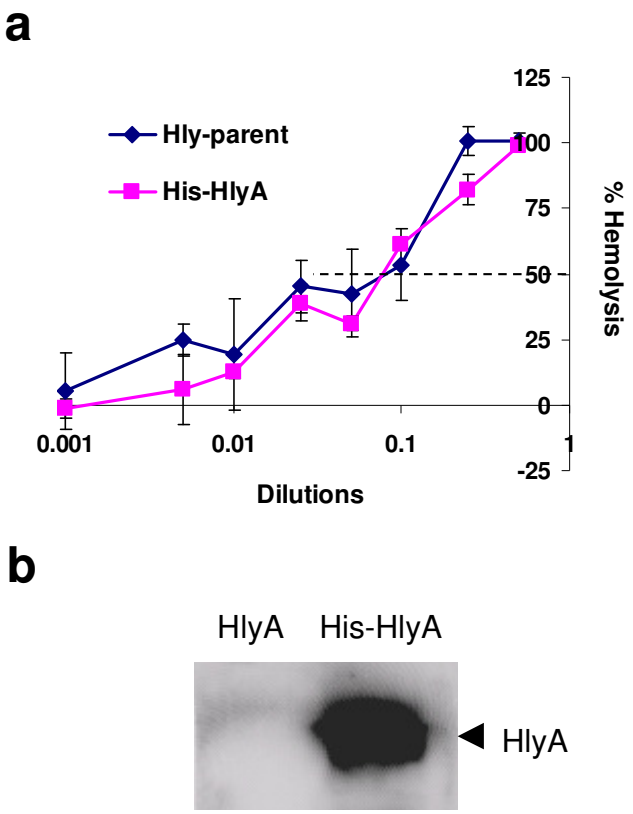


Figure 6.8. 6X-His tagged HlyA has similar activity as HlyA as quantified by liquid blood lysis assay (**a**). The supernatant dilution corresponding to 50% hemolysis was calculated (numbers in bold marked by blue dashed arrows) from the lysis curve and secretion fold change was calculated by taking the ratio of the dilutions. (**b**) Western blot analysis of the supernatant fraction indicates that mouse anti-6X-His antibody (Sigma, MO, USA) interacts preferentially with the 6X-His tag on HlyA protein and not with secreted HlyA itself.

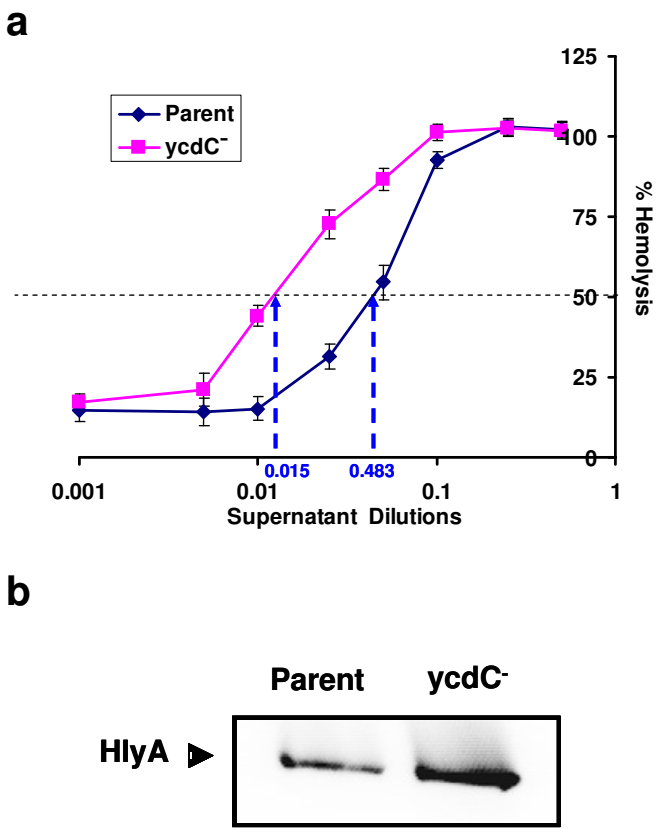


Figure 6.9. (a) Liquid blood lysis assay to quantify the amount of secreted active HlyA protein in Hly-ycdC^{+} and Hly-ycdC^{-} strains. Each reading was performed in triplicates and error bars show the standard deviation of triplicate measurements. The supernatant dilution corresponding to 50% hemolysis was calculated (numbers in bold marked by blue dashed arrows) from the lysis curve and secretion fold change was calculated by taking the ratio of the dilutions. (b) The 3.4-fold increase in HlyA secretion was also verified by normalized western blot analysis of the secreted HlyA protein.

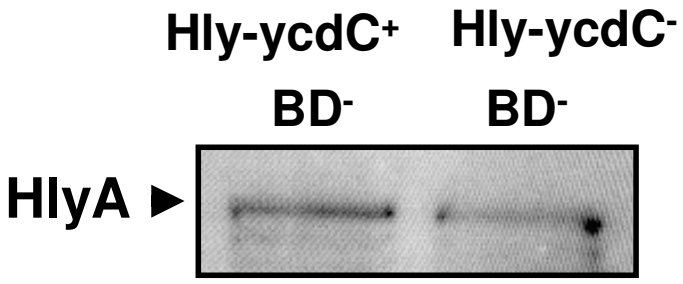


Figure 6.10. Normalized western blot analysis of intracellular HlyA protein in the secretion deficient strains, Hly-ycdC^{+} (BD^{-}) and Hly-ycdC^{-} (BD^{-}). Secretion deficient strains were used to decouple the effects of secretion and intracellular protein production. Samples were normalized on the basis of cell density and quantitative analysis of the western blots was done using ImageMaster 2D Platinum Software v5.0 (GE Amersham Biosciences, NJ, USA).

background. Western blot analysis of the supernatant fractions indicates a 42% increase in Bla secretion in the Bla-ycdC$^-$ relative to Bla-ycdC$^+$ strain (Fig. 6.11). Intracellular quantitation of Bla in secretion deficient cells revealed a 30% decreased expression in the Bla-ycdC$^-$ (BD$^-$) strain relative to Bla-ycdC$^+$ (BD$^-$) strain (Figure 6.11). We previously observed that decreased HlyA expression can result in active HlyA hypersecretion [14], and it has been hypothesized that a significant increase in protein expression can overwhelm the secretory pathway [18]. Hence, we hypothesized that the absence of functional RutR (coded by *ycdC* gene) may increase active HlyA secretion by decreasing the HlyA translation rate and promoting a balance between HlyA translation and secretion.

To further investigate the effect of the absence of functional RutR on HlyA translation rate, mRNA expression profiling of Hly-ycdC$^-$ and Hly-ycdC$^+$ strains was performed using Affymetrix GeneChip® *E. coli* Genome 2.0 array. Gene expression analysis, using GeneSpring™ software (Agilent, CA, USA), detected 240 genes that are differentially expressed in the two strains. The data are presented as absolute ratios of the normalized gene expression between the Hly-ycdC$^-$ and Hly-ycdC$^+$ strains. Values greater than 1.0 indicate that the gene was expressed more highly in the Hly-ycdC$^-$ mutant and values less than 1.0 indicate a lower expression level in the Hly-ycdC$^-$ mutant. Expression of the carbamoylphosphate synthase (CarAB) small subunit gene (*carA*) was decreased 2.36-fold in the Hly-ycdC$^-$ mutant relative to the Hly-ycdC$^+$ strain (Table 6.2), consistent with a recent study which showed that RutR is a positive regulator of the *carAB* operon [19]. CarAB utilizes glutamine as the natural amino group donor and provides carbamoylphosphate for arginine and pyrimidine biosynthesis [20,21]. Arginine biosynthesis was shown to be interlinked with proline biosynthesis via early intermediates in the two pathways [22]. Thus, decreased *carA*

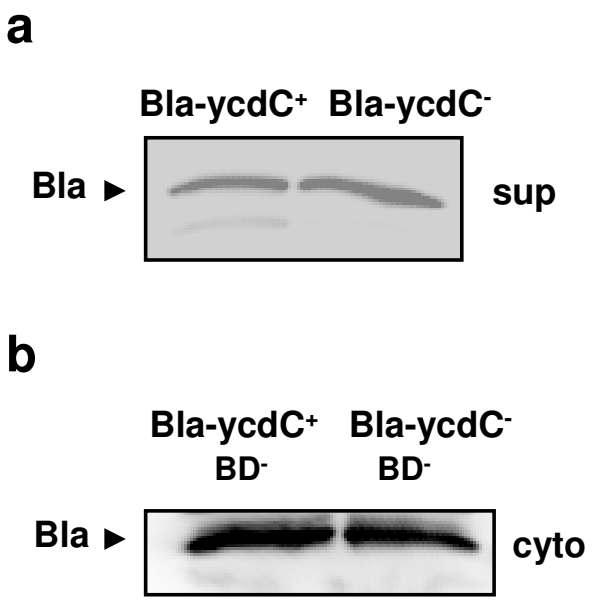


Figure 6.11. Absence of a functional ycdC increases the secretion of beta-lactamase (Bla) protein. Normalized western blot of (**a**) extracellular Bla expression (sup) in ycdC^{+} and ycdC^{-}strains, and (**b**) intracellular Bla expression (cyto) in secretion deficient ycdC^{+} (BD^{-}) and ycdC^{-} (BD^{-}) strains. Samples were normalized on the basis of cell density and quantitative analysis of the western blots was done using ImageMaster 2D Platinum Software v5.0 (GE Amersham Biosciences, NJ, USA).

Table 6.2. mRNA fold change levels of carA and amino acid transporter genes for proline, arginine, tyrosine and histidine in the Hly-ycdC$^-$ mutant relative to the Hly-ycdC$^+$ strain. Values less than 1.0 indicate a lower expression level in the hly-ycdC$^-$ mutant and vice versa.

Gene Name	**Gene Title**	**Fold change (Hly-ycdC$^-$ / Hly-ycdC$^+$)**
carA	carbamoyl phosphate synthase small subunit	0.42
proY	proline permease transport protein	0.82
putP	Proline:sodium symporter	0.87
artJ	arginine transporter subunit	0.71
tyrP	tyrosine-specific transport protein	0.61
hisM	histidine transport system membrane protein M	0.62
hisQ	histidine transport system permease protein hisQ	0.69

expression might have a cascade effect on these biosynthetic pathways resulting in decreased intracellular levels of arginine and proline. We also observed a decreased expression of amino acid transporter genes for arginine, proline, tyrosine and histidine in the Hly- ycdC$^-$ mutant (Table 6.2), which suggests a decreased import of the respective amino acids present in the extracellular media. Also, the mRNA levels of tRNA-synthetase genes were found to be either decreased or unchanged in the Hly-ycdC$^-$ strain relative to the Hly-ycdC$^+$ strain (Table 6.3), consistent with our previous observation in B41 strain[1]. This genome-wide analysis reveals that RutR protein play an important role in maintaining the balance of various substrates (amino acids, tRNA-synthetases) of the *E. coli* translation machinery and the absence of functional RutR may result in slower translation rates of proteins, and in particular recombinant HlyA. This premise is consistent with our observation that there is a decreased intracellular HlyA and Bla expression in secretion deficient strains (Fig. 6.10 and 6.11).

6.6 Conclusion

In this study, we report the discovery of a single nucleotide polymorphism (G → T) in the hypersecreter B41 genome relative to the Hly-parent genome. Different factors that may account for the effect of the SNP on HlyA secretion phenotype were investigated and our data suggests that the absence of functional RutR resulted in decreased intracellular recombinant protein expression. Gene expression analysis revealed that absence of RutR coordinates a decrease in the expression of *carA*, tRNA-synthetases and some amino acid transporter genes. Although, these results suggest that the absence of functional RutR promotes the balance between translation and secretion by slowing the translation rate, the mechanism needs to be further investigated. Overall, these studies present a new target for metabolic engineering to enhance extracellular secretion via the Type-I pathway. More generally, the results

Table 6.3. mRNA fold change levels of tRNA-synthetase genes for proline, arginine, tyrosine and histidine in the Hly-ycdC$^-$ mutant relative to the Hly-ycdC$^+$ strain. Values less than 1.0 indicate a lower expression level in the hly-ycdC$^-$ mutant and vice versa.

tRNA-synthetase (gene name)	Fold change (Hly-ycdC$^-$ / Hly-ycdC$^+$)
alaS	0.84
argS	0.88
asnS	0.97
aspS	0.96
cysS	0.76
glnS	0.68
gltX	0.83
glyQ	0.72
glyS	0.60
hisS	0.72
ileS	0.64
leuS	0.71
lysS	0.86
lysU	0.98
metG	0.66
pheS	0.83
pheT	0.86
proS	0.58
serS	1.01
thrS	0.79
trpS	0.94
tyrS	0.84
valS	0.68

reported here demonstrate the utility of short read genome sequencing technology to identify gene targets and genetic variations that can accelerate the development of novel organisms for biotechnology applications.

6.7 Acknowledgements

The authors would like to thank Peter Schweitzer and James VanEe at the Cornell University Life Sciences Core Laboratories Center for performing the Illumina® genome sequencing, Frederick Blattner for sharing *ycdC::*Tn5 strain (Strain # FB22271 from *E. coli* K-12 MG1655 Genome Initiative) and Rodney Welch for kindly donating hemolysin plasmids. K.H.L. is supported by NYSTAR, NSF, and University of Delaware.

REFERENCES

1. Lee PS, Lee KH. 2005. Engineering HlyA Hypersecretion in *Escherichia coli* Based on Proteomic and Microarray Analyses. *Biotechnol. Bioeng.* **89**: 195-205.

2. Loh KD *et al*. 2006. A previously undescribed pathway for pyrimidine catabolism. *Proc. Natl. Acad. Sci. USA* **103**: 5114-5119.

3. Ley TJ *et al.* 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**: 66-72.

4. Maher C. *et al.* 2009.Transcriptome sequencing to detect gene fusions in cancer. *Nature* doi:10.1038/nature07638.

5. Holt KE *et al.* 2008. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella Typhi*. *Nat. Genet.* **40**: 987-993.

6. Crossman, L. *et al.* 2008. The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biol.* **9**: R74.

7. Thomas RK *et al.* 2007. High-throughput oncogene mutation profiling in human cancer. *Nat. Genet.* **39**: 347–351.

8. Dahl F *et al.* 2007. Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc. Natl. Acad. Sci. USA* **104**: 9387–9392.

9. Craig, D.W. *et al.* 2008. Identification of genetic variants using bar-coded multiplexed sequencing. *Nat. Methods* **5**: 887-893.

10. Velicer GJ *et al.* 2006. Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. *Proc. Natl. Acad. Sci. USA* **103**: 8107–8112.

11. Andries K *et al.* 2005. A Diarylquinoline Drug Active on the ATP Synthase of *Mycobacterium tuberculosis. Science* **307**: 223-227.

12. Felmlee T, Welch RA. 1988. Alterations of amino acid repeats in the *Escherichia coli* hemolysin affect cytolytic activity and secretion. *Proc Natl Acad Sci USA* **85**: 5269-5273.

13. Thomas WD Jr, Wagner SP, Welch RA. 1992. A heterologous membrane protein domain fused to the C-terminal ATP-binding domain of HlyB can export *Escherichia coli* hemolysin. *J Bacteriol*. **174**: 6771-6779.

14. Gupta P, Lee KH. 2008. Silent mutations result in HlyA hypersecretion by reducing intracellular HlyA protein aggregates. *Biotechnol. Bioeng.* **101**: 967-974.

15. Pimenta AL, Young J, Holland IB, Blight MA 1999. Antibody analysis of the localisation, expression and stability of HlyD, the MFP component of the *E. coli* haemolysin translocator. *Mol. Gen. Genet.* **261**: 122-132.

16. Vakharia H, German GJ, Misra R. 2001. Isolation and characterization of *Escherichia coli* tolC mutants defective in secreting enzymatically active alpha-hemolysin. *J. Bacteriol.* **183**: 6908–6916.

17. Jurgens D, Ozel M, Takaisi-Kikuni NB. 2002. Production and characterization of *Escherichia coli* enterohemolysin and its effects on the structure of erythrocyte membranes. *Cell Biol. Intl.* **26**: 175-186.

18. Simmons LC, Yansura DG. 1996. Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. *Nat. Biotechnol.* **14**: 629-634.

19. Shimada T, Hirao K, Kori A, Yamamoto K, Ishihama A. 2007. RutR is the uracil/thymine-sensing master regulator of a set of genes for synthesis and degradation of pyrimidines. *Mol. Microbiol.* **66**: 744-757.

20. Piérard A, Glansdorff N, Mergeay M, Wiame JM . 1965. Control of the biosynthesis of carbamoyl phosphate in *Escherichia coli*. *J. Mol. Biol.* **14**: 23-36.

21. Piérard A, Wiame JM 1964. Regulation and mutation affecting a glutamine dependent formation of carbamyl phosphate in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **15**: 76-81.

22. Berg CM, Rossi JJ. 1974. Proline excretion and indirect suppression in *Escherichia coli* and *Salmonella typhimurium*. *J. Bacteriol.* **118**: 928-939..

CHAPTER 7

CONCLUSION AND FUTURE DIRECTIONS

7.1 Summary

The first part of this thesis highlighted the use of 'directed silent mutagenesis' to enhance recombinant protein secretion in *Escherichia coli*. This method incorporates synonymous rare codon clusters at specific sites in the target genes, based on the predictions by a previously developed mathematical model of translation [1]. A significant increase in the secretion of active HlyA protein via the Type-I pathway was observed and the effect of a synonymous rare codon cluster on HlyA secretion was studied in detail. Different factors (including mRNA expression, secretion machinery expression, protein degradation) were investigated that may account for the enhanced secretion phenotype and the data suggested that the observed phenotype is the result of decreased total HlyA protein production relative to the hly-parent strain. It was also shown that most of the intracellular HlyA exists in the inclusion body fraction. The results illustrated that production of high levels of secreted proteins appears to require a balance between translation and secretion rate. Synonymous rare codon engineering enhanced secretion of recombinant proteins not only via the Type-I pathway but also via other pathways (Sec, Tat, and HasABC).

The secretion pathways were also used to study the effects of synonymous substitutions on protein folding. Type-I secretion studies suggested that the effects of synonymous mutations are protein specific and can affect the quality control mechanisms (protein aggregation, degradation) that are triggered by the overexpression of a recombinant protein. The studies of protein secretion via the Tat and the HasABC exporters strongly suggested that the presence of abnormal

translation kinetics, caused by the ribosome moving slower through the rare codon cluster, can alter the final protein conformation and activity. It was also observed that synonymous rare codon substitutions can alter the interaction of a polypeptide with an unfolding/folding modulator. The association of a molecular chaperone (EF-Tu) with beta-lactamase protein tagged for secretion via the Type-I and the Tat pathways was discovered and it was hypothesized that EF-Tu might impede secretion by preventing the polypeptide to fold in the secretion-competent conformation. Based on these observations, the presence of secretion-competent and secretion-incompetent pools of polypeptides was suggested. It was also hypothesized that the introduction of rare codons in a specific stretch of the target gene can affect the balance of these pools and the secretion flux, since failure to rapidly reach a secretion competent conformation may either result in partial or complete deposition into insoluble aggregates or degradation. These results offer insights into the possible mechanisms of how silent mutations can affect protein structure and function.

The second part of the thesis illustrated the use of 'next-generation' sequencing technology to profile the hypersecreter mutants, created by random mutagenesis. This work involved sequencing a previously isolated derivative hypersecreter *E. coli* strain (B41) [2] and parent strain using Illumina® sequencing technology. Mutational profiling of B41 and parent genomes revealed a single nucleotide polymorphism (G → T) in B41 genome, which results in premature translation termination of a transcription factor (RutR). Different factors that may account for the effect of the SNP on HlyA secretion phenotype were investigated and the data suggested that the absence of functional RutR resulted in decreased intracellular recombinant protein expression. Comparative gene expression analysis was also performed which revealed that absence of RutR coordinates a decrease in the expression of *carA*, tRNA-

synthetases and some amino acid transporter genes. Although, these results suggest that the absence of functional RutR promotes the balance between translation and secretion by slowing the translation rate, the mechanism needs to be further investigated. These studies presented a single gene target to enhance extracellular secretion via the Type-I pathway and highlighted the potential of new high-throughput massively parallel sequencing technologies to characterize selected mutants for strain improvement.

7.2 Recommendations for future work

7.2.1 Quantitative understanding of the affect of synonymous rare codon cluster on protein secretion

To predict the effect of synonymous rare codon cluster on protein secretion, it is important to understand the relationship between the type of rare codon used and the position of the rare codon cluster. Hence experiments need to be designed in such a way to decouple the two parameters and study their effect individually on secretion yield. I have observed from the experiments, with beta-lactamase (Bla) as the model protein, that there is a position-dependent effect of the incorporation of a rare codon cluster. In particular, we observed an increased secretion of Bla in a mutant strain containing the synonymous rare codon cluster near the 3' end of the gene. This observation might be because of Bla folding change resulting in more secretion compatible protein. Further experiments can be performed to understand the codon specific effects of an interesting mutant (having an increased or a decreased secretion phenotype relative to the parent strain). In other words, different combinations of codon changes for a particular mutant of the model gene can be made and tested for secretion and intracellular protein expression. This kind of combinatorial approach will give some clues about the combinations of codon changes which are most

important for affecting the secretion yield for a particular position of the rare codon cluster. Similar experiments can be done for other model proteins and the data generated from these sets of experiments can be fed into the secretion model to obtain estimates on the rate parameters for protein aggregation and translation rate to better understand and predict the secretion behavior of various mutants for a given recombinant protein. This strategy would also help in designing mutant strains more objectively for a given protein.

7.2.2 Study the effect of chaperone proteins on protein secretion

In recombinant cells, proteolysis and aggregation can compete with each other and can be observed as antagonistic events for folding-reluctant proteins [3]. Since folding, aggregation and degradation are all competing processes, the protein processing step is very important in deciding the fate of the polypeptide, targeted for secretion. Both chaperones and proteases are components of the quality control system in the cell [4], devoted to surveying the folding status of cellular proteins. But it is not clear how the discrimination between folding attempts or proteolysis occurs [5] or if specific signals in target polypeptides are involved in deciding between both events. It has been reported in the literature that overexpression of aggregated protein can be minimized by controlling process parameters such as temperature, reducing recombinant gene expression, protein engineering or by the co-expression of plasmid encoded chaperone genes [3]. To minimize the amount of aggregate formation in the system and direct more protein towards secretion, the effect of chaperones on the secretion capacity can be explored. Recently, it has been argued that GroEL is involved in both the protein removal from the inclusion bodies as well as in the promotion of inclusion body formation in bacteria by clustering of small aggregated nuclei [6]. Hence it will be interesting to explore the effects of over-expression and deletion of some key

chaperones on the secretion of model proteins. Secretion studies of a "slow" mutant and parent strain for at least two model proteins can be performed in two types of strains; one with a GroEL(-) background and one in which GroEL and GroES are being co-expressed on a plasmid vector. The GroEL(-) strain can be created using standard gene knockout techniques (e.g. homologous recombination). For further analysis, similar studies on the DnaK chaperone can be performed.

7.2.3 Creation of synonymous codon library

I observed from my studies that the presence of a synonymous rare codon cluster can affect the folding of the polypeptide. This observation adds evidence to the growing literature that silent mutations can change protein structure and function [7,8,9] and challenges the general notion that protein evolution occurs only by changes in the primary amino acid sequence. In order to further validate that protein folding may be nucleotide sequence dependent, a synonymous rare codon library of a reporter protein (e.g. green fluorescent protein) can be created, which would consist of gene sequences having different nucleotides but encoding the same amino acid sequence. This may be achieved by using semi-randomized weighted oligonucleotide synthesis and end-to-end ligation [10]. The library will be semi-randomized because only the third base of the codons will be randomized in order to achieve the same amino acid sequence. The synonymous library can also be fused with an N-terminal signal sequence (ssTorA) for secretion via the Tat pathway. This experiment can be used to detect the "super-folders" and the "poor-folders" because the Tat pathway has an inherent quality control mechanism and secretes only the folded polypeptides [11,12]. Once an interesting mutant is discovered, biophysical characterization can be performed to link the changes in the gene sequence to the changes in the secondary structure.

7.2.4 Exploring the mechanism of RutR protein for enhanced secretion

The studies involving genome sequence comparison of B41 and parent genome revealed that the absence of functional RutR can result in enhanced secretion of recombinant proteins via the Type-I secretion pathway. The study hypothesized that RutR might affect the balance between different substrates of the translation machinery in *E. coli*, thereby affecting the translation rate of all proteins and in particular, the recombinant protein of interest. An intracellular free-amino acid quantitation can be performed to test this hypothesis. Additionally, the data indicates decreased expression of *carA* gene in the B41 strain, consistent with the idea that CarAB protein might also be an important player in the RutR mediated translation regulation. Hence, secretion and intracellular protein studies can be performed in a *carA* deletion mutant. Also, RutR effect and synonymous rare codon cluster effect may be simultaneously studied by transforming hly-slow and other mutants into the ycdC deletion background. The expectation is that the combined effect may slow down the translation of the recombinant protein even further and may result in higher secretion yields.

7.3 Conclusion

The work presented here includes different type of strategies to improve protein secretion in *E. coli*. A directed silent mutagenesis approach illustrates the utility of synonymous rare codon engineering to improve recombinant protein secretion via multiple secretion pathways for bioprocess applications. These studies also reveal that the incorporation of synonymous rare codon clusters at specific sites can modulate the interactions with specific molecular chaperones and can drive the *in vivo* folding of the polypeptide chain into different conformations. This research not only opens up a new avenue to study the effect of silent mutations on protein folding/misfolding, but also

suggest that silent single nucleotide polymorphisms (SNPs) should not be neglected in determining the likelihood of the development and progression of many diseases such as Alzheimer's disease and others that are strongly linked to SNPs [13]. I also used 'next-generation' sequencing technologies in tandem with genomic expression analysis to characterize selected mutants leading to successful metabolic engineering strategies for strain improvement. These studies not only presents a single gene target in *E. coli* to enhance the extracellular secretion of recombinant proteins, but also illustrate the promise of genome sequencing towards rational design of interesting phenotypes for biotechnology applications.

REFERENCES

1. Lee PS, Lee KH. 2005. Engineering HlyA Hypersecretion in *Escherichia coli* Based on Proteomic and Microarray Analyses. *Biotechnol. Bioeng.* **89**: 195-205.

2. Shaw L, Zia R, Lee KH. 2003. Totally asymmetric exclusion process with extended objects: A model for protein synthesis. *Phys. Rev. E* **68**: 021910.

3. Villaverde A, Carrio MM. 2003. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol. Lett.* **25:**1385-1395.

4. Villaverde A, Carrio MM. 2003. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol. Lett.* **25**: 1385-1395.

5. Herman C, D'Ari R. 1998. Proteolysis and chaperones: the destruction/reconstruction dilemma. *Curr. Opin. Microbiol*. **1**: 204–209.

6. Carrió MM, Villaverde A. 2003. Role of molecular chaperones in inclusion body formation. *FEBS Lett.* **537**: 215–221.

7. Kimchi-Sarfaty C. *et al*. 2007. A''silent'' polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**: 525-528.

8. Komar AA, Lesnik T, Reiss C. 2002. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett.* **462**: 387–391.

9. Cortazzo P. *et al.* 2002. Silent mutations affect in vivo protein folding in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **293**: 537-541.

10. Isalan M. 2006. Construction of semi- randomized gene libraries with weighted oligonucleotide synthesis and PCR. *Nat. Protoc.* **1**: 468-75.

11. Sanders C, Wethkamp N, Lill H. 2001. Transport of cytochrome c derivatives by the bacterial Tat protein translocation system. *Mol. Microbiol.* **41**: 241–246.

12. DeLisa MP, Tullman D, Georgiou G. 2003. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc. Natl. Acad. Sci.* **100**: 6115–6120.

13. Emahazion T *et al.* 2001. SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trend Genet.* **17**: 407-413.

APPENDIX A

A.1 Effect of temperature on HlyA secretion by the Type-I pathway

We observed from our studies that a decreased intracellular HlyA aggregate formation, as a result of synonymous rare codon cluster, can result in enhanced active HlyA secretion. Previous studies have observed that lowering the culture temperature can slow down the translation rate. Hence, we tested if a decrease in culture temperature can affect the secretion yield of active HlyA protein. HlyA secreting cells were grown at 37 °C and 30 °C. Supernatant fractions were analyzed for active HlyA using liquid blood lysis assay. The results indicated that cells grown at 37 °C secrete nearly 2.6-fold more active HlyA relative to the cells grown at 30 °C (Fig. A.1). This observation indicates that lower temperature might exert some other effects on the Type-I secretion system and needs to be further evaluated.

A.2. Positional effects of the rare codon cluster on Bla secretion via the Sec and Tat pathways

Positional effects of the synonymous rare codon cluster on Bla secretion by the Sec and Tat pathways were investigated. The *de novo* gene sequence (Fig. 5.7) was used for these experiments. Synonymous mutants are synthesized by replacing the new codon cluster with the 'rare' codon cluster (GGG ATA CTA CTA) at each position (Table A.1). Quantitative western analysis of the periplasmic fractions indicated that Sec strains secrete significantly less GILL-Bla protein (Fig. A.2). We also observed that there is a significant decrease in the cytoplasmic GILL-Bla expression in the Sec-M3 mutants (Fig. A.2). The observations suggest that GILL-Bla protein expressed with a Sec signal peptide attains a conformation, which is secretion inhibitory. For the Tat pathway, quantitative western analysis indicated a 2-fold increase in secretion in

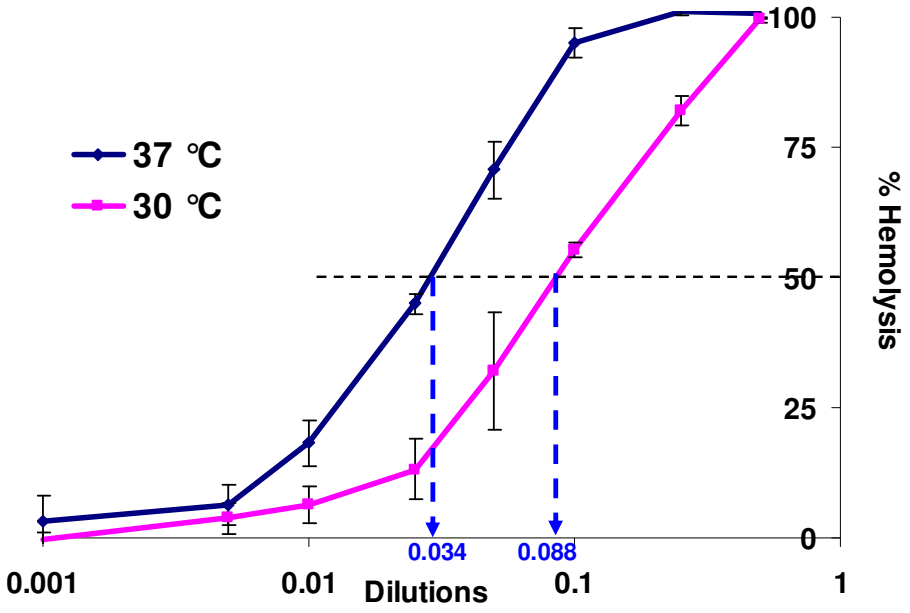


Figure A.1. Liquid blood lysis assay to quantify the amount of secreted active HlyA protein in the cells grown at 37 °C and 30 °C. Each reading was performed in triplicates and error bars show the standard deviation of triplicate measurements. The supernatant dilution corresponding to 50% hemolysis was calculated (numbers in bold marked by blue dashed arrows) from the lysis curve and secretion fold change was calculated by taking the ratio of the dilutions.

Table A.1. Sequence changes (marked in red) and predicted % decrease in translation rate of different sec-bla and tat-bla mutants

(b) GILL-Sec-Bla Mutants	Sequence	% predicted Slowdown
gillsecM1	210 AAAGGTA TCCTGCTGTG T initial sequence AAAGGGA TACTACTATG T rare codon seq. 70 K G I L L C amino acid seq.	5.3%
gillsecM2	402 AACGGTA TCCTGCTGAC A initial sequence AACGGGA TACTACTAAC A rare codon seq. 134 N G I L L T amino acid seq.	19%
gillsecM3	726 CGCGGTA TCCTGCTGGC A initial sequence CGCGGGA TACTACTAGC A rare codon seq. 242 R G I L L A amino acid seq.	10.6%

(b) GILL-Tat-Bla Mutants	Sequence	% predicted Slowdown
gilltatM1	297 AAAGGTA TCCTGCTGTG T initial sequence AAAGGGA TACTACTATG T rare codon seq. 99 K G I L L C amino acid seq.	5.3%
gilltatM2	558 AACGGTA TCCTGCTGAC A initial sequence AACGGGA TACTACTAAC A rare codon seq. 186 N G I L L T amino acid seq.	19%
gilltatM3	882 CGCGGTA TCCTGCTGGC A initial sequence CGCGGGA TACTACTAGC A rare codon seq. 294 R G I L L A amino acid seq.	10.6%

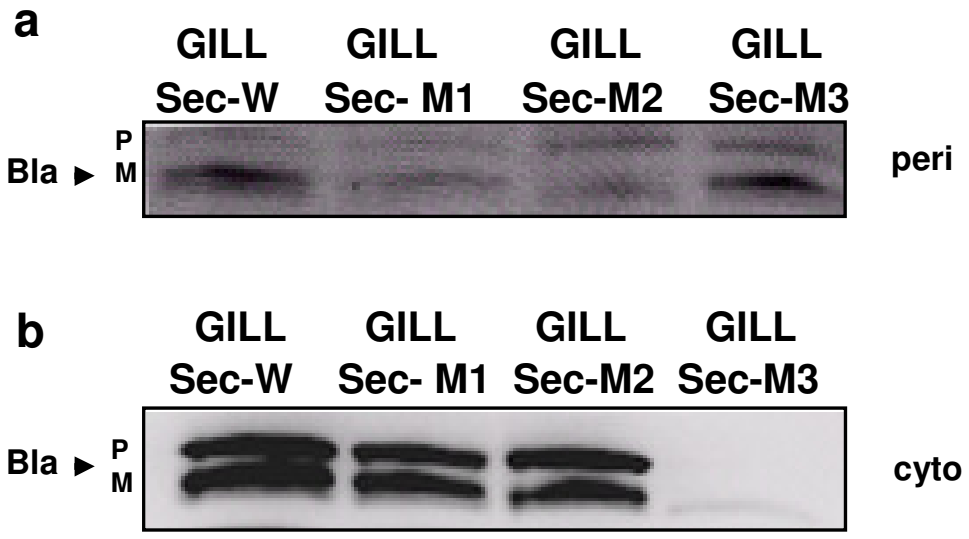


Figure A.2. Positional effect of rare codon cluster on Bla secretion by the Sec pathway. Normalized Western analysis of Bla in (a) periplasmic fraction (peri), and (b) cytoplasmic fraction (cyto). An equivalent number of cells were harvested. P & M are precursor and mature form of Bla.

GILL-Tat-M1 strain relative to GILL-Tat-W parent strain (Fig. A.3). However, significant decrease in secretion was observed for the other two strains relative to the parent strain. The GILL-Tat-M3 strain has a 60% decrease in GILL-Bla intracellular expression (Fig. A.3).The results suggest that the presence of the synonymous rare codon cluster near the 5'end can facilitate GILL-Bla secretion via the Tat pathway. However, the cluster near the 3'end severely affects its ability to fold into a secretion competent conformation.

A.3. Bla secretion by the Sec and Tat pathways in a *ycdC*::Tn5 background

It was observed from our experiments that there is an enhanced secretion of α-hemolysin and Bla protein by the Type-I pathway in *ycdC*::Tn5 background. The effect of a non-functional RutR on Bla secretion by the Sec and the Tat pathways was also investigated. The results of Bla secretion by the Sec pathway indicated no bla secretion in the periplasmic fraction of both the parent and ycdC$^-$ strains (Fig. A.4). Faint Bla bands were observed in the cytoplasmic fractions of the two strains, suggesting that either the intracellular Bla was aggregated or degraded (Fig. A.4). A relatively dark band was observed in both the periplasmic and cytoplasmic fractions at around 55 kDa, suggesting the presence of Bla dimers (Fig. A.4). The absence of Bla secretion by the Sec pathway may be a strain dependent effect, since MG1655 strain was used in these experiments. Bla protein was to secrete relatively well by the Tat pathway in both the parent and ycdC$^-$ strains; however there was no relative difference in the band intensities (Fig. A.5). Moreover, most of the Bla protein present in the periplasmic fraction contained precursor Bla. No difference was observed in the cytoplasmic fractions of the two strains as well (Fig. A.5).

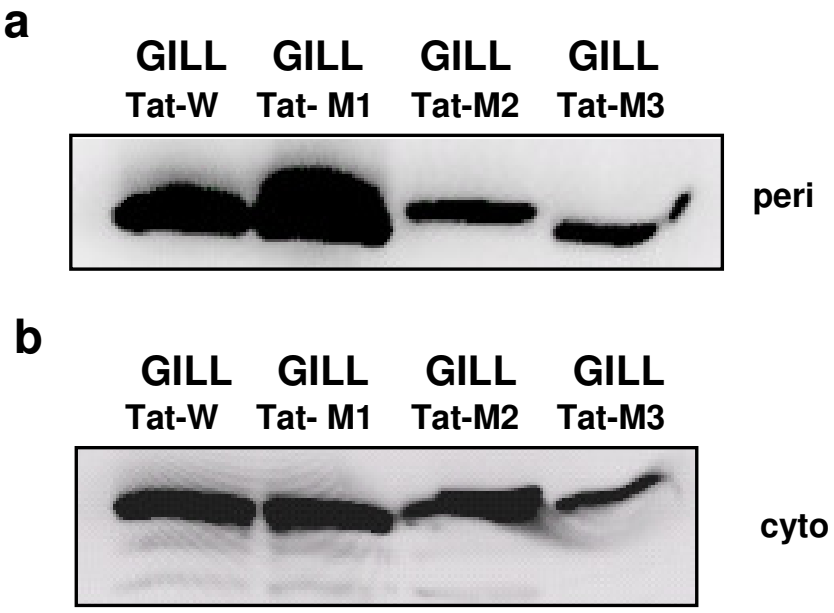


Figure A.3. Positional effect of rare codon cluster on Bla secretion by the Tat pathway. Normalized Western analysis of Bla in (a) periplasmic fraction (peri), and (b) cytoplasmic fraction (cyto). An equivalent number of cells were harvested.

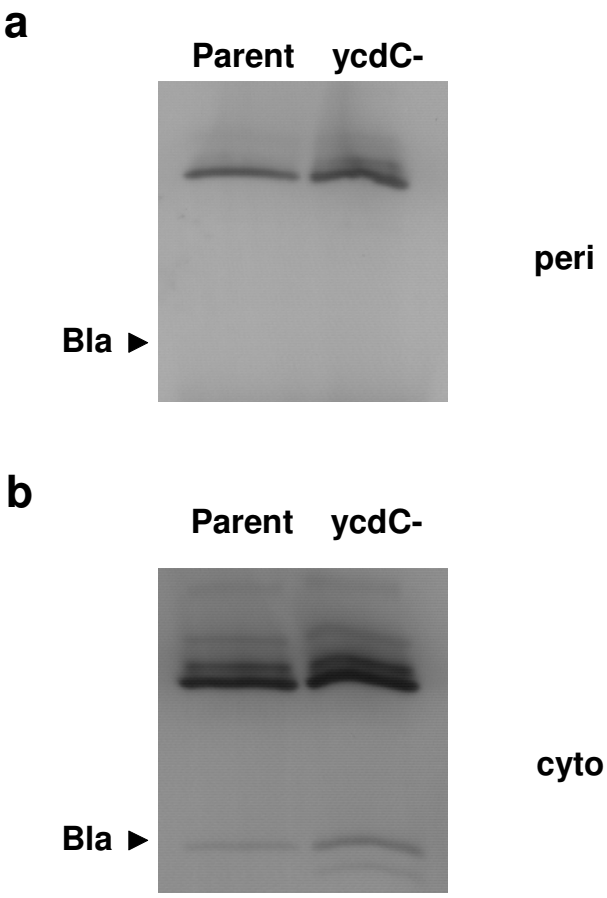


Figure A.4. Bla secretion by the Sec pathway in *ycdC*::Tn5 background. Normalized Western analysis of Bla in (a) periplasmic fraction (peri), and (b) cytoplasmic fraction (cyto). An equivalent number of cells were harvested

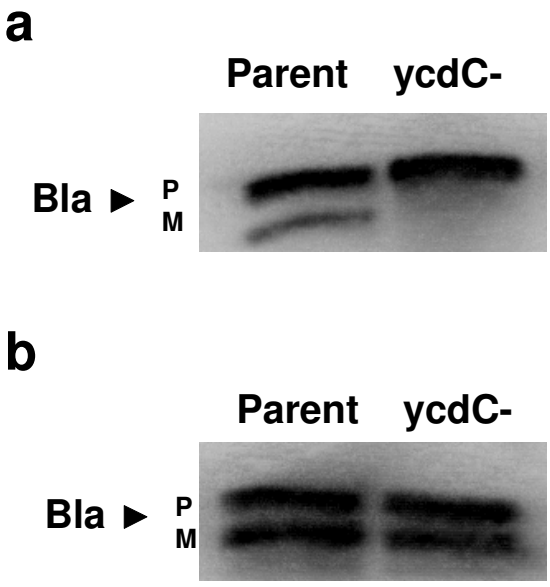


Figure A.5. Bla secretion by the Tat pathway in *ycdC*::Tn5 background. Normalized Western analysis of Bla in (a) periplasmic fraction, and (b) cytoplasmic fraction. An equivalent number of cells were harvested. P & M are precursor and mature form of Bla.